UNIVERSITY OF
LINCOLN

EVALUATING HUMAN DECISIONS DURING TIME AND ACTOR INDEPENDENT VARIABLES.

Charles Walker

School of Computer Science

College of Science

University of Lincoln

Submitted in partial satisfaction of the requirements for the
Degree of Master of Science by Research
in Computer Science

Supervisor Dr. Mark Doughty and Dr. Chris Headleand

January 2020

1. Acknowledgements

I would like to thank the University of Lincoln ICT Department for their help and understanding during the research degree. Further to this, I am profoundly grateful to Matthew Cavill for giving me the time and authorisation to do my degree whilst working Full-Time.

A special thank you to Wayne Gathergood for dealing with me on the Train when talking about my Degree and providing advice about the Masters process.

I would also like to thank my close family for their support and understanding throughout my two-years as a Masters student.

I would also like to thank Phil Assheton for his assistance in understanding the data coming out of both studies. Without him, the analysis of the results would not be the same.

Finally, this Thesis would not have been possible without the supervision of Mark Doughty, Chris Headleand and Antonella De Angeli.

2. Abstract

The Trolley Problem, a well known thought experiment comparing decisions during life or death circumstances, was applied by the Moral Machine. This gained 40 million decisions from millions of participants. Whilst accepted and praised for its success, investigation is available of participants position in the environment and the effect of time-pressure on the decisions made.

To answer this, a web study was conducted to gain a quantitative understanding of participants likelihood to make life or death decisions under the effect of the independent variables via generalised estimating equation. The effects of which proved to be non-significant across both independent variables. Time-pressure showed self-sacrifice to be twice as likely when under time-pressure ($B = 0.512$, $p = 0.012$). This effect was studied via a quantitative and qualitative virtual reality study, understanding whether the significance is repeatable. The results indicate the opposite, showing regardless of the independent variable, participants are likely to sacrifice themselves. The explanation of the prior studies findings being concluded as 5% false positive in regards to significance.

The implications of both studies provide validation into the Moral Machine's results, showing the independent variables not chosen by the Moral Machine had little significance on participants decisions. This provides understanding around the development of a Trolley Problem algorithm in autonomous vehicles and the effects that would occur in the world. The research also provides a recommendation that research is required to understand the time taken to make a decision during both time and non-time pressure decisions. This would be to see if non-time pressure is being treated as such.

Table of Contents

List of Figures

List of Tables

3. Thesis Structure

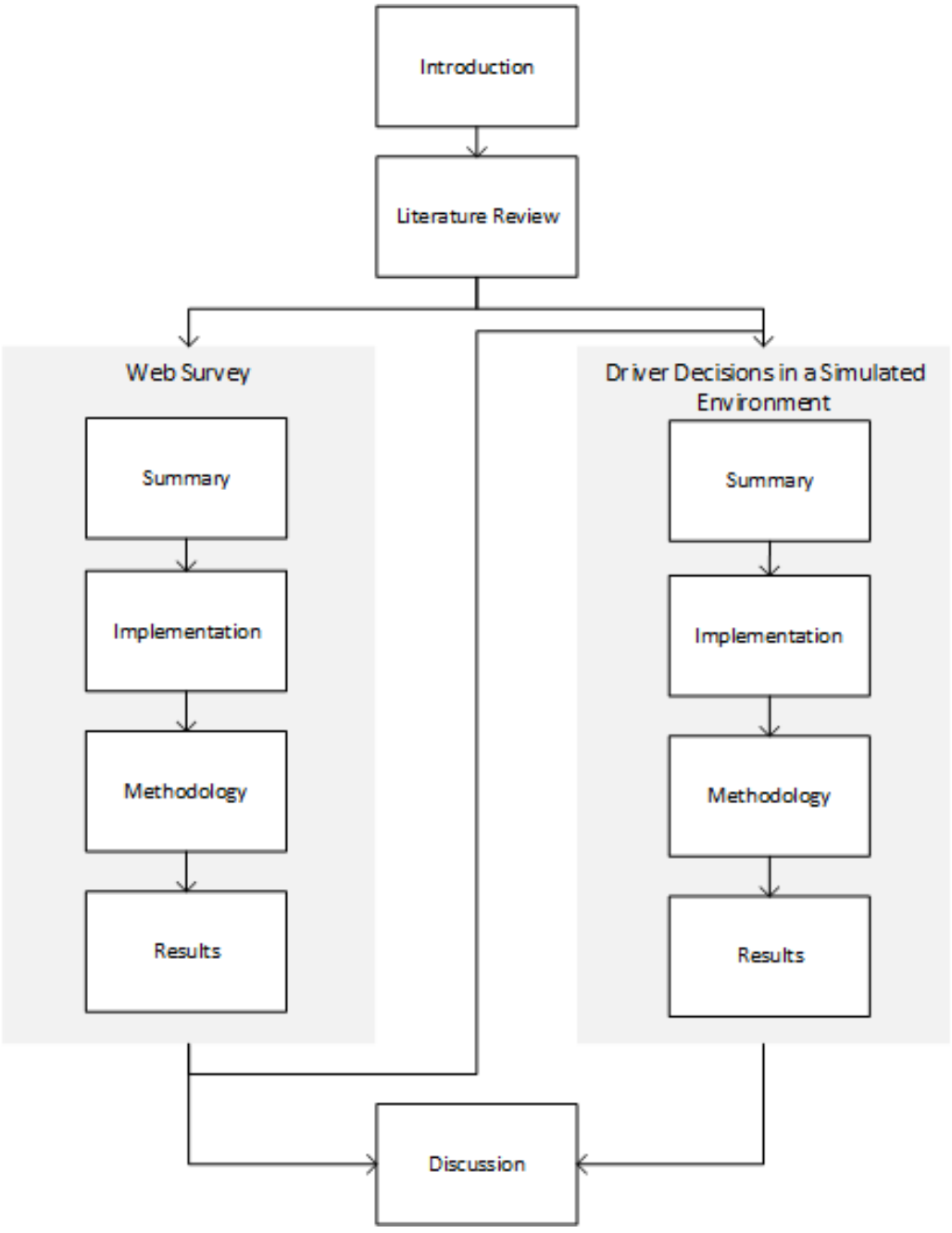

Figure 3.1: Structure of the thesis chapters.

4. Introduction

Personal transportation technology has dramatically changed over the years. With the development of cruise control, assisted parking, emergency braking assistance and now semi-autonomous cars (Jardine Motors Group, 2019), there is little doubt that technology in vehicles is increasing at a substantial rate.

With that said, automated cars are still an emerging technology where semi-autonomous vehicles have only recently become commercially available, like Tesla's self driving technology, which does not fully take the control away from the driver, expecting them to keep their hands on the wheel, and always remain vigilant (The Tesla Team, 2015). Whereas fully autonomous cars, are still several years away, with Knapman from the Telegraph predicting that autonomous cars can be used in most circumstances by 2025 (Knapman, 2016). Questions around liability are the focus of attention in the news, which to sum up is being answered with: "Assigning liability depends on what action led to the collision and whether it was based on decisions by the driver or the vehicle." (Jurdak and Kanhere, 2018)

Research also questions what people think automated vehicles should do when the human is fully removed from the driving equation. This is where this thesis stems. The Moral Machine, "an online experimental platform designed to explore the moral dilemmas faced by autonomous vehicles." (Awad et al., 2018, p. 59) and currently one of the most successful ethical studies with regards to participant count with "millions of people in 233 countries and territories" having "logged 40 million decisions" identified "that participants from individualistic cultures, like the UK and US, placed a stronger emphasis on sparing more lives given all the other choices" (Hao, 2018). The Moral Machine presented a pair-wise comparison which always resulted in some form of catastrophe, be it bigger or smaller than the alternative option. These choices came with an unlimited amount of time to decide which side the vehicle was going to travel down. This raised the question whether this could

cause an unrealistic expectation on the automated vehicle, when humans who are involved in collisions have seconds to decide.

This is the first hypothesis:

> There is a significant difference in decisions between time and non-time sensitive collision scenarios.

The Trolley Problem, a thought problem first introduced by Foot, thinking through the consequences of an action and the determination of its value based on the outcome (D'Olimpio, 2016), was further elaborated on by Thomson with regards to potential decision differences between where a human is placed in a scenario (Thomson, 1985).

This leads onto the final hypothesis:

> There is a significant difference in collision decisions when participants are placed in different areas of an environment.

Both these research questions in this thesis were tested using an online web survey which aimed to follow the Moral Machine's footsteps in visual design, but utilise both independent variables. To then gain further insight into the time-pressured independent variable, a comparison of self-preservation to self-sacrifice decisions were evaluated in a virtual reality environment to provide a qualitative understanding of the decisions participants were making, rather than the prior study which was fully quantitative.

4.1 Motivation

The hypotheses presented and evaluated within this document, provides two interesting outcomes dependent on whether they are proven or not.

Should time-pressure be proven to have a significant difference in participants decisions, it will show that more understanding is required of the Trolley Problem with regards to autonomous ethics. The reason for this, is that the Moral Machine, having only evaluated non-time pressured collisions, could cause an unrealistic expectation

on automotive manufacturers to implement a solution that could subsequently cause autonomous vehicle users to feel more uncomfortable about a collision, if they were to be involved in one.

When evaluating the actor independent variable, should it be presented as significant, it must be questioned as to what this could mean for the Moral Machine's results should it be viewed as an implementable option. The reason for this, is that the Moral Machine only asked what the driver should do and did not provide other environmental locations for the participant.

On the flip-side of both hypotheses, should Time-Pressure or Actor be proven to be non-significant, further validation to the Moral Machine's success into gaining participants ethical choices would be presented. It would therefore be possible to recommend the decision choices that the Moral Machine presents as a possible solution in the event of a Trolley dilemma collision scenario.

4.2 Order of Information in the Thesis

To begin with, this thesis will identify, compile and explain existing research that has occurred around autonomous vehicles with regards to ethics and user viewpoints about the technology.

The review shows how the research questions were formed. The thesis then describes the implementation of both the web survey and the virtual reality environment, explaining how the use of specific tools, frameworks and programming languages allowed the implementation of the independent variables and study constraints. Due to the implementation of both studies being somewhat similar, the implementation has been concatenated together.

The thesis then splits the studies down into separate sections, isolating their explanation of relevance, methodologies and results. The reason for this was that both studies can be treated in isolation. This is because for both studies data was primary

collected and only within the discussion section do the results of both studies combine to form an overall answer to the research question.

Within each study, the methodology explains what types of data were gathered, how the data was analysed, and the sample design used. From here, the results of the studies is presented, with key areas of significance and insignificance shown.

Specifically in the virtual reality study, thematic analysis results are presented within their key themes, whilst further separated into sub-themes.

After the results of both studies are presented, a discussion of the results is available, providing interpretations of the results and highlighting if the hypotheses were true whilst explaining the connection with the literature review. Further to this, the implications of the results are presented, to explain what the results could mean outside of the thesis. Limitations of the thesis are presented, explaining why certain areas cannot be answered from this thesis. Within the discussion section, there is finally a recommendation section, explaining what actions should be taken to best utilise the results found from the thesis, as well as a list of possible research areas that could be investigated to provide further explanation to areas of ethics in autonomous systems.

This leads onto the conclusion section which wraps up the thesis by explaining the impact the thesis could have, as well as the areas of research others could consider venturing into from this study.

5. Literature Review

The area of automation in vehicles is being rapidly researched due to the improvements in vehicle communication technology like Vehicle to Vehicle Communication and Vehicle to Infrastructure Communication (House of Commons Library, 2017, p. 3) as well as technology centred around Human-Vehicle Communication, for example, understanding human behaviour around a vehicle, thus hoping to improve vehicle understanding of pedestrian intent (Ohn-Bar and Trivedi, 2016, p. 95). As we move towards vehicles with higher autonomy we open "new research avenues in dealing with learning, modelling, active control, perception of dynamic events, and novel architectures for distributed cognitive systems. Furthermore, these challenges must be addressed in a safety-time critical context" (Ohn-Bar and Trivedi, 2016, p. 100). Perception of dynamic events is important to the argument of this research. How would people react to dynamic events?

5.1 Social Attitudes Towards Autonomous Vehicles

At the centre of the argument, the current understanding of existing social attitudes and prior acceptability of autonomous vehicles, indicates a correlation between acceptability of a new technology and the attitudes of an individual (Payre, Cestac and Delhomme, 2014, p. 253). The research consisted of three phases of studies, two pilot studies and one main, large scale study. In the two pilot studies, a range of questions were asked to participants relating to their views on autonomous vehicles. The results showed an overall acceptance, but also indicated boundaries that consumers felt would make them more comfortable using the technology, some of which included using "automated driving for long journeys" or refusing to "use such a device in a city". On the flip side, the study also revealed that two out of five participants would be willing to have an autonomous car drive for them when they were under

the influence of alcohol or suffering from side-effects of medication (Payre, Cestac and Delhomme, 2014, p. 255).

This is a worrying outcome, despite the two out of five not being a majority value, that is still a potentially large proportion of people who could be intoxicated within an autonomous vehicle that may or may not encounter an issue when under operation. However, this study states that for this pilot study it only recruited five participants and it is therefore not possible to imply that the sample size is enough to effect a broader population.

Another study states that “an automated driving system will allow the driver to take his eyes off the road and engage in non-driving related tasks”. This was demonstrated within a driving simulator, which was dependent on visual and physical fidelity "drivers adapt compensating behaviors that allow for realistic responses but may not fully reflect how the driver would respond in the real world" (Philips and Morton, 2015, p. 10). The results showed that drivers are willing to do so, possibly increasing the demand of a take-over situation (Körber, Baseler and Bengler, 2018, p. 19), which is a common testing theme when evaluating trust in autonomous vehicles. If the results of the previously mentioned study are taken within the context of this study, it is possible to see a potential issue; consumers are beginning to see autonomous vehicles as an entire substitution of the driver from the driving system (Payre, Cestac and Delhomme, 2014, p. 253).

In a study of 149 participants, there were situations that put the participant in a VR environment, whereby the study gauged the users guilt level via extracting information from the forum by explicitly looking for feelings of guilt or not after an extreme ethical situation(Cristofari and Guitton, 2014, p. 2). VR environments and measuring guilt were chosen because “guilt has been consistently reported as an important emotion in the development of moral insights" (Cristofari and Guitton, 2014, p. 5) whilst VR has “been demonstrated to display stronger emotional reactions in response to virtual reality rather than text" as well as "reactions to virtual persons have been found to be similar to reactions to people in real life" (Cristofari and Guitton, 2014, p. 1). From the study, 120 of 149 situations showed evidence of guilt,

whilst 29 situations were coded as not guilty (Cristofari and Guitton, 2014, p. 3). Further to this, 149 situations were classed as being self-justified, which meant the participant did think the decision was correct. The most interesting outcome of the study was the fact that "actions with immediate consequences caused more guilt than actions with delayed consequences" (Cristofari and Guitton, 2014, p. 5). This highlights an interesting point which may also be observed in the studies shown in this document. Due to the outcome of the situation being immediate, it might be possible to observe a similar outcome.

Byrne brings an interesting argument to the table, stating that driving is not "just a mechanical operation but also a complex social activity". He states that cars could be produced so that they follow well-defined rules: "If obstacle and traveling fast: swerve. Else: stop. If gap: merge", however he goes onto say that the mechanics of a car "involve subtle interactions between humans that reflect those of the not-driving world." (Bryne, M. 2017) This was identified by Brown who utilised YouTube videos of autonomous cars to gain an understanding of their actions in real-world conditions. This totalled around ten and a half hours of footage from around the world (Brown, 2017, p. 92). In the YouTube videos, "most of the time autopilot drives without incident. Yet, due to its simple mechanics, autopilot sometimes misunderstands other drivers' actions" (Brown, 2017, p. 93). Brown goes onto provide examples of this occurring, one of which being a two-lane highway. The autonomous vehicle is offered to overtake another vehicle in the slower lane by a silver car in the faster lane. Due to the silver car in the fast-lane being present, the autonomous vehicle refused to go into the lane, "the silver car's driver understandably perceives this as doubly rude" (Brown, 2017, p. 94). The examples Brown provides can indicate a weakness in autonomous vehicles in relation to their way of handling social situations, one of which could be applied to the handling of collisions and even the "pre-actions" (Brown, 2017, p. 95) given beforehand. He does however state that from the YouTube videos collected, "most of the time autopilot drives without incident." This is useful because it shows autonomous cars are safer but the pitfall of lack of social understanding will cause multiple incidents until "all vehicles are fully

autonomous" (Brown, 2017, p. 92) or social understanding is correctly implemented into the vehicles.

5.2 The Trolley Problem

The Trolley Problem was proposed by Foot (Foot, 1967, p. 2), and further elaborated on by Thomson; they propose a concept where a trolley is approaching five men on a track. Due to a fault the trolley is unable to stop, and the trolley driver must decide whether to kill the five men, or turn the trolley into the siding where there is one person (Thomson, 1985, p. 1397). When Thomson delves further into the matter, she questions areas of the Trolley Problem previously not looked at by Foot. Firstly, Thomson investigates whether different real-world problems provide a different outcome by looking at the scenario where a surgeon has a choice to operate on one man, transplanting his vital organs into five patients. This would save the five patients but kill the one. Alternatively, the surgeon chooses to spare the one patient, which would result in the death of the five patients. Based on the outcome she gave, she determined that sacrificing the one was worse than letting five die (Thomson, 1985, p. 1396). This is derived from how a surgeon is sworn to cause no harm to another individual. By killing the one person to save the five, that surgeon would be going against their oath and so shows that the Trolley Problem is only a small segment of ethical models. Thus context can provide a wide range of different factors that are not initially considered. Where this decision differs for the trolley driver, comes from the fact that due to the driver being liable for the safety of the people around the trolley and in the trolley (Thomson, 1985, p. 1397), he has a different decision to make, thus meaning that by killing the one person and saving the five, it could be viewed that he has caused less harm, via the adoption of the Utilitarian perspective; the concept where saving the majority over the minority of people is considered to be justified.

Thomson then moves to a different dilemma based around the Trolley Problem, this time involving someone stood by the side of the tracks, with a lever which can change

the direction of the trolley. The trolley driver, for some reason, is incapacitated. This means the person by the tracks is given the decision. Where this differs from the trolley driver making the decision, is that rather than the trolley driver making two active decisions to kill one or five; the person by the tracks makes either a passive or an active decision. The person by the tracks could choose to make no decision, thus being completely oblivious to the outcome, nor the fact that that person has no gain or loss from the decision (Thomson, 1985, p. 1397). This indicates that the social circumstances and environmental perceptions of the decision maker is then to be considered when evaluating the decision.

Thomson's final elaboration of the Trolley Problem moves back to the trolley driver and now poses a different critique on the outcome that could happen from killing one: the social background of the five men are unknown. "The five are not track workmen at all, but Mafia members in workmen's clothing, and they have tied the one work-man to the right-hand track in the hope that you would turn the trolley onto him" (Thomson, 1985, p. 1398). This provides even further elaboration into Foot's initial thought experiment.

5.3 Implementations of the Trolley Problem

From the prior issues such as drivers not being present due to not focusing on the road, completely disengaged in other tasks and the lack of social handling from autonomous vehicles, this is clearly a growing and pressing matter that may subsequently lead to further issues. For example, when an autonomous vehicle enters an environment where a collision is unavoidable and people either in the environment, or present within the vehicle will become injured, or worse killed. This issue was investigated by the MIT Moral Machine which chooses to incorporate the theory of the Trolley Problem. The study involved participants selecting between a pair of images showing the outcome of an automated vehicle collision. One showing the collision on the left side of the road, and another on the right side of the road. With the collection of 40 million decisions came the analysis of results via pair-wise com-

parison between the images using conjoint analysis, to identify social preferences in collision scenarios against pre-configured demographics.

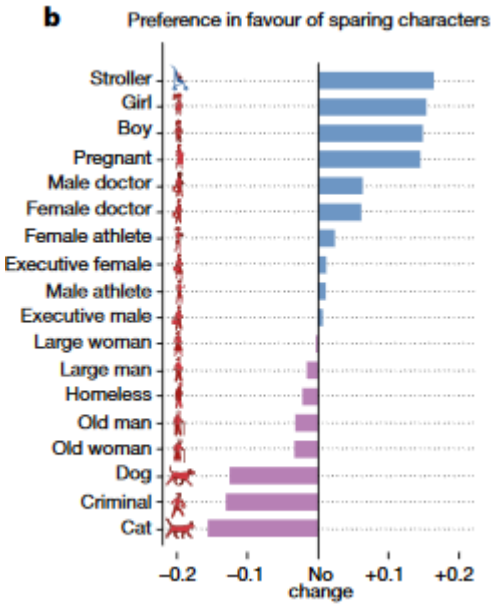

Figure 5.1: Moral Machine: Preference in favour of sparing characters

The results indicated that the "strongest preferences are observed for sparing humans over animals, sparing more lives, and sparing young lives" (Awad et al., 2018, p. 60).

However, the Moral Machine has some shortcomings as reported by Reese, with having no element of time-constraint it did not simulate people's reactions in a realistic setting thus "Humans in a moment of panic are rarely equipped to make moralistic decisions to choose between killing one or two people," (Reese, 2016) which could have dramatically changed the outcome of the Moral Machine results. A viewpoint from Reese is that when "polling people about moral decisions, while the results may be intriguing, it is not my idea of how to give engineers basic material for programming a self-driving car's moral decisions." Whilst this could well be the case, the Moral Machine researchers did not state the direct application of the model towards real-life autonomous vehicles and instead was designed to "contribute to developing global, socially acceptable principles for machine ethics" (Awad et al., 2018, p. 1). Arguably, this does indicate some intent towards changing the way machine ethics reacts to those situations. A further shortcoming, as described by Nyholm, is the lack of research that is related to gaining justification about someone's decisions.

This is important because “in ethical arguments, it is important to articulate and assess arguments in favour of or against the different options that are being considered” (Nyholm, 2018, p. 5). Nyholm is further sceptical of the Moral Machine because “most people still have very little real experience with self-driving cars. It is likely that their attitudes will change once they have more actual experience with them. This suggests that we should not put too much weight on people’s current attitudes about this technology” (Nyholm, 2018, p. 5). Whilst there is certainly a point with regards to not putting too much weight on people’s current attitudes, it is also important that gaining a baseline could be vital in later understanding what the impact of autonomous vehicles have on the general populace.

An observation that can also be made about the Moral Machine is the lack of anti-robot mechanisms on the website. It could well be that an automated mechanism could have added a multitude of records into the Moral Machine’s statistics which would have potentially created skewed results from what could have been gathered (Basso and Miraglia, 2008, p. 149).

There have been other studies implementing the Trolley Problem. In a study involving 62 law students at the University of Eastern Piedmont in Alessandria in Italy, participants were requested to complete a questionnaire during one of two days around two scenarios:

The lever-pulling scenario, which consisted of "A passer-by" who "could pull a lever next to the track, and this way deviate the trolley onto the side-track. The passer-by realises that, if he does not pull the lever, the five people will be killed. If he pulls the lever instead, the five people will be saved. The passer-by is aware, however, that by pulling the lever the person on the side-track will be killed" (Lanteri, Chelini and Rizzello, 2008, p. 795).

This is followed by an overweight stranger scenario, consisting of a passer-by standing "next to the track, and he could push a very fat stranger onto the trolley’s path, halting its ride. The passer-by realises that, if he does not push the stranger, the five people will be killed. If he pushes the stranger instead, the five people will be

saved. The passer-by is aware, however, that by pushing him, the stranger will be killed" (Lanteri, Chelini and Rizzello, 2008, p. 795).

Participants were shown one of the two scenarios in an alternating fashion, then seeing reversed occasions of the scenarios. It was evident that the overweight human scenario in both participant groups responded in the majority that it was immoral to sacrifice the overweight human. The most interesting outcome was the comparison between the reversal of the scenarios, "when the lever scenario is put second, fewer participants are willing to operate on the switch than when it is put first, but the responses to the stranger scenario remain unaffected" (Lanteri, Chelini and Rizzello, 2008, p. 796). This indicates that responses to the lever scenario are crucially affected by the order that scenarios are shown, whilst showing that "emotional activation of the stranger scenario makes participants more alert to personal moral violations" (Lanteri, Chelini and Rizzello, 2008, p. 797). This was further explained due to prior understanding of both scenarios; the pushing of the stranger is intentional. In contrast, the lever scenario is viewed as an impersonal decision.

These results were similarly found in another study which showed five different scenarios to fifty participants, with the scenarios being ordered in two ways; Least Agreeable First and Most Agreeable First. From these orderings, results indicated that when participants were shown Least Agreeable scenarios first, they were more likely to continue this trend by viewing other scenarios similarly, whereas Most Agreeable showed a trend that would gradually decline as the agree-ability decreased (Wiegmann, Okan and Nagel, 2012, p. 822). This mirrors the prior study, showing that emotional effect could well have been at play and that moral reasoning was changed due to prior experience.

5.4 Virtual Reality and Ethical Dilemmas

Whilst the use of web surveys can be beneficial to gain large amounts of quantitative data, like that of the Moral Machine; there is proof that immersion via VR can have an influence on the decisions being made by participants. A study was conducted to

identify this via a within-subject, order-dependent experiment which involved both text-based and VR dilemmas. The study recruited forty participants between the ages of 18 and 28 (Patil et al., 2014, p. 64) and with them completing both the text-based and VR environments, led to a result which indicated differences between the text-based and VR dilemmas "with many of them behaving in utilitarian manner in VR dilemmas despite their non-utilitarian judgments for the same dilemmas in textual descriptions" (Patil et al., 2014, p. 94). This was due to text-based judgements having a utilitarian outcome of around 0.76; in comparison the VR session was 0.95, "therefore, the difference between the proportions of utilitarian decisions taken in the two sessions was significant" (Patil et al., 2014, p. 100).

This idea of immersive VR environments was further studied in an environment dedicated to travelling back through time and changing the course of history. The idea being, to see "whether the ability to go back through time, and intervene, to possibly avoid all deaths, has an impact on how the participant views such moral dilemmas, and also whether this experience leads to a re-evaluation of past unfortunate events in their own lives" (Friedman et al., 2014, p. 1). To try and gauge the level of immersion that a participant was feeling, they evaluated three specific types of illusion; presence, body presence and agency with presence scoring a median subjective level of 6 as well as body presence scoring a median of 5 which was "well in line with previous studies" (Friedman et al., 2014, p. 9). It is important to note that this study does not directly compare traditional survey methods against VR, however the use of the three illusion types, is an indication of what can be done to ensure the environment is as immersive as possible.

One study chose to focus on specific influences in the Trolley Problem, compared to the Moral Machine which provided a huge variety of combinations. The study involving sixty-six participants looked at decisions between gender, ethnicity, body orientation and quantity, removing other factors (Skulmowski et al., 2014, p. 4). In addition to this, the study implemented an independent variable of music to try and identify if music could change emotional responses to the scenarios. From the study, the most prominent outcome was that of quantity which indicated that 96% of the

time, participants would sacrifice the one to save the many. Music, in the case of these results showed no significant effect (Skulmowski et al., 2014, p. 7).

What is interesting about this study, is that it used time pressure to force participants to decide. From the ANOVA comparison, it was possible to identify that ethnicity, gender and body orientation had slower response times than group comparisons, demonstrating an interesting effect that group decisions were easier to make than other comparisons (Skulmowski et al., 2014, p. 8). Another study showed similar results, finding that 95.4% of 189 participants chose to sacrifice the single person over the side of the road with more living obstacles (Bergmann et al., 2018, p. 5).

One study compared time pressure where pressure is induced by the amount of time until a collision, in comparison to the prior study which evaluated reaction time from a constant time limit. This study used a car, rather than a trolley and pitched participants against a range of demographics, such as the Moral Machine. When comparing the variable time pressure, from the fast condition, the participant error increased "four-fold" "from the slow condition" (Sütfeld et al., 2017, p. 9). Interestingly, the higher the time pressure, the less likely a participant was to sacrifice a male adult over a female one, which results from prior comparisons across the same time-pressure value had yielded. The results "speculated tendency toward social desirability" and "would likely rely on slower cognitive processes, and thus not come into effect in fast-paced intuitive decisions" (Sütfeld et al., 2017, p. 10). This finding demonstrates that there could be more understanding needed to truly gauge whether time pressure causes different choices. Although this study did provide a variable level of time-pressure, it did not then evaluate whether no-time pressure had any further impact on the comparison; this is something that is important to understand before conclusions are drawn on automated ethical systems.

An alternative to the "classic Trolley Problem" (Bergmann et al., 2018, p. 6) was tested, trying to identify whether self-preservation influenced how participants would react. Would participants save themselves, a single person, or a range of people? When participants were presented with the option of killing themselves or two others, 52% of the time they chose themselves. (This is not significantly different however

was identified as being more altruistic than the study had envisaged (Bergmann et al., 2018, p. 6).) This self-sacrifice percentage increased as the number of people in the road did, with the result of seven people in the road achieving 70% self-sacrifice rate. These results further show a utilitarian manner towards ethical dilemmas, which consistently seems to be the option across most studies.

5.4.1 The Validity of Virtual Reality Ethical Studies

Questions still exist surrounding the validity of ethical studies via VR. “Kantian duty ethics, first of all, upholds as the most fundamental moral principle that human beings have a duty to treat other persons with respect.” “However, a virtual person is not by any measure a real person but is merely a simulation of a person” (Brey, 1999, p. 8). As stated by Brey, this concept needs empirical evidence and as such should be treated as inconclusive. It is an interesting idea to contemplate however, as stated by Parsons, “virtual reality environments proffer assessment paradigms that combine the experimental control of laboratory measures with emotionally engaging background narratives” (Parsons, 2015, p. 1). This means that VR does provide a compromise and allows safe testing of physical situations, in comparison to conducting studies in real-life.

5.5 Alternatives to the Trolley Problem

An alternative to the Trolley Problem is known as the Tunnel Problem, originally created by Millar (Millar, 2014). Whilst like the Trolley Problem, the Tunnel Problem encompasses the vehicle with a tunnel, addressing arguments around the Trolley Problem not being scalable to vehicles because of the "infinite" possibilities a car could run into, unlike a trolley which is on tracks (Technative, 2018). Instead, the application of a tunnel reduces the option set down to a similar binary format. This is achieved by the vehicle being enclosed in an environment that only allows the deviation onto different lanes.

Whilst there is the question of what the car should do, the question the tunnel problem tries to answer is who should decide what the car should do?

The argument made by Technative and Millar is that the owner of the vehicle should be able to decide how their car reacts to such a situation. Regardless, Technative continues to state that "AI technologies are going to have to continue to adapt" to new situations regardless of how many ethical dilemma solutions are implemented.

5.6 Existing Collisions from Autonomous Vehicles

This idea of collision decision making is something that has become a more pressing matter recently due to a few collisions that have occurred with Self-Driving Cars and other road users or pedestrians. The most notable was a collision between a Self-Driving Uber and an Arizona Woman (Levin and Wong, 2018). When Levin interviewed Simpson, the privacy and technology project director with Consumer Watchdog stated “the robot cars cannot accurately predict human behaviour, and the real problem comes in the interaction between humans and the robot vehicles” (Levin and Wong, 2018). Further reports of the incident say that the vehicles “sensors detected Herzberg” however were “tuned too far in favour of ignoring objects in its path which might be “false positives” (such as plastic bags)” (Gibbs, 2018). This can also be elaborated because “most successful algorithms still have remarkably low success rates when identifying cyclists”, “even when the weather is good” (Renda, 2018, p. 3). This identifies a need for further research and development into Human-Robot Interaction when based in the field, as well as an improvement in algorithms used during the detection of objects as well as a consideration about single-points of failure on a vehicle and the sensors used to detect those around the vehicle.

Uber’s vehicles haven’t been the first to be involved in a collision. The recorded first incident was Tesla when the vehicles sensors were impaired by a “bright spring sky”. This caused the vehicle to collide into the back of an 18-wheeled truck crossing the highway, impacting the windshield and causing the fatality of the self-driving vehicles occupant (Yadron and Tynan, 2016). This collision does not necessarily suggest that

the collision decision making technology needs improvement, but instead questions why the sensors did not report an impairment, and then perform necessary actions; further highlighting the need for better Human-Robot Interaction.

Around two years prior to the Uber's collision, improvements were researched within new prototypes of automated vehicles. Via the use of information such as "lane and road information" they "can be combined with pedestrian detection and tracking for performing intent-aware path prediction and activity classification". To achieve a factor of this, the assessment of "body pose, and head pose can be used to infer pedestrian intent to cross and predict paths." When combined with map information, this provides a level of risk estimation of "pedestrians around a vehicle." (Ohn-Bar and Trivedi, 2016, p. 95). It is however unknown whether these algorithms were implemented in new models of automated vehicles therefore it is difficult to classify if the algorithm could have, or failed to negotiate the risk the Uber Vehicle was faced with.

Automated vehicles are not the only area of automation that have been shown to be at risk of causing accidents; not only physically, but socially, culturally and politically (Crawford and Calo, 2016, p. 311). One example was when Google "tweaked its image-recognition algorithm in 2015 after the system mislabelled an African American couple as gorillas." In response to the issue, Google also proposed introducing a 'red button' into its AI systems should the system get out of control. This issue has not been the only reported inaccurate justifications; in some contexts "AI systems disproportionately affect groups that are already disadvantaged by factors such as race, gender and socio-economic background". One example if this, being that an investigation in 2016 yielded information that proprietary algorithms widely used by judges to help determine the risk of re-offending are almost twice as likely to mistakenly flag black defendants than white defendants. As well as Google's search engine, in 2013, there would have been a twenty-five percent higher chance to flag up advertisements for criminal-records when querying "names commonly used by black people" compared to "white-identifying names" (Crawford and Calo, 2016, p. 312). With the evidence provided above, it does raise the question about how vehicles could treat those within groups that are classed as disadvantaged. How-

ever, the hope is that several mitigations are now being implemented. For example, the German Federal Ministry of Transport and Digital Infrastructure published a report stating, “Distinction between any Human stature e.g. age is strictly prohibited” (Federal Ministry of Transport and Digital Infrastructure, 2017, p. 2010) which hopefully will provide a level of recommendation to vehicle manufacturers to implement this standard. Further to this, Crawford and Calo provide evidence of what companies have done since the prior issues, an example being firms deploying frameworks (such as value sensitive design) to help them identify likely stakeholders and their values. With that said, "the concern remains that corporations are relatively free to field test their AI systems on the public without sustained research on medium- or even near-term effects" (Crawford and Calo, 2016, p. 312).

5.7 Who to Blame after a Collision?

With the implementation of ethical decision systems, and the understanding of human viewpoints on those dilemmas, another factor can be raised, “Our laws are ill-equipped to deal with autonomous vehicles” (Fournier, 2016, p. 42). This is an interesting point that is further backed up by the German Federal Ministry of Transport and Digital Infrastructure which released a report stating guidelines on ethical decisions within autonomous decisions. These guidelines state that the responsibility of the ethical decision implementation falls to the manufacturer (Federal Ministry of Transport and Digital Infrastructure, 2017, p. 10) whilst considering that there is a priority the vehicle should follow when faced with a collision. This involves humans always being put above everything else, with animals and property being the second thing to preserve, so long as no human is harmed (Federal Ministry of Transport and Digital Infrastructure, 2017, p. 10). At first glance, this may seem as though this is in-fact counteracting the argument from Fournier; it is however, due to guidelines not being legally binding documents and having only been released within the past few months. There may be some delay until implementation and because the committee was based in Germany, not all vehicle manufacturers may adopt the same viewpoint. To further argue the lack of legal preparation, “the UK Government has not begun

to address these issues of 'algorithmic morality'" (House of Commons Library, 2017, p. 10), although a committee in the House of Lords for science and technology discussed the matter, giving different viewpoints. Some believing the implementation of algorithmic morality is "a good thing for road safety", whilst others considering that it is not "achievable or desirable" (House of Commons Library, 2017, p. 10). The conflict in opinions is expected in this source, due to it being a government discussion. Regardless, these are all steps towards legally binding rules, but at this moment in time, leaves the algorithms open to interpretation.

5.8 Implementation Options of Ethical Systems

5.8.1 Personal Ethics Settings

Whilst there are legal arguments towards who would be responsible, there are questions as to the implementation of ethics algorithms and who decides on these. One argument is the implementation of a Personal Ethics Setting (PES) providing owners the ability to decide how the car should react in an ethical dilemma, also known as an "ethical knob" (Contissa, Lagioia and Sartor, 2017, p. 377). Whilst this could provide drivers with the comfort of knowing what the vehicle will do, this form of implementation would most likely cause a prisoner's dilemma, the understanding of "what governs the balance between cooperation and competition"(Dixit and Nalebuff, 2019), causing a selfish PES, thus a higher rate of competition (Gogoll and Müller, 2017, p. 698). This could be negated by a series of disincentives, for example, higher insurance premiums or limited insurance coverage (Contissa, Lagioia and Sartor, 2017, p. 378). PES would answer the legal question of who is responsible. Should PES be the implementable option, there needs to be laws in place that "should determine what level of user-selected egoism could lead to an AV behaviour that could expose the user to criminal or civil liability" (Contissa, et al. 2017, 378).

5.8.2 Mandatory Ethics Settings

The alternative option is to use a Mandatory Ethics Setting (MES) which would be a decision from a third party that would affect all vehicles, providing a consensus of ethical outcome. Gogoll and Müller argue "that people would not be willing to use an automated car that might sacrifice themselves in a dilemma situation", however further state that "MES is in the considered interest of everybody" (Gogoll and Müller, 2017, p. 698) and recommend that an MES is designed to minimise overall harm. Whilst minimisation is a valid argument, there is also the question of how a vehicle evaluates minimisation of harm. If there is an inevitable crash and the vehicle must choose between hitting a pedestrian or a motorcyclist, the concept of harm minimisation would mean that the vehicle would target the motorcyclist due to the motorcyclist wearing protective equipment. This is discrimination. Perhaps if the motorcyclist was not wearing any protective gear, the vehicle would choose to hit the pedestrian due to the motorcyclist now being more at risk due to added momentum from the motorcycle (Lin, 2016, p. 73).

Hevelke and Nide-Rümelin argue the idea of "strict liability" meaning that those who choose to own and use an autonomous car should be collectively held accountable for the outcomes, which could be covered by "a tax or a mandatory insurance". They further argue that autonomous vehicles would save lives, therefore companies should be encouraged to continue without the development being too risky for a company to undertake, whilst also maintaining standards should a development be unsuitable, such that rectification is completed in a timely manner (Hevelke and Nida-Rümelin, 2015, p. 629). This could be a middle-ground between PES and MES, whilst also following existing vehicle driving standards.

This is where the argument of legal boundaries is most important. There needs to be laws to provide the companies with the necessary boundaries of their developments, whilst also putting to bed the numerous vast questions around ethics, justice and discrimination (Schreurs and Steuwer, 2015, p. 168).

Goodall poses an alternative solution, around risk management, which would provide

a more everyday solution to risks faced whilst driving. The methods it utilises is based on trying to foresee outcomes of a decision and allocating the severity of the risk for each outcome (Goodall, 2016, p. 814). Goodall states "an automated vehicle needs a way to determine if the benefits of moving into the left lane outweigh the costs" (Goodall, 2016, p. 815). The use of risk management is a feasible, (and already well under way) improvement in vehicle automation, but what it does not answer at this current time, is how people would feel when a collision was to occur, should the risk management system either fail, or decide the risk of having to collide with one object was less than another option. This method seems to aim at not requiring ethical collision systems at all, mitigating any argument around why a collision was decided; instead aiming at reducing the risk of collisions occurring at all. Goodall does acknowledge that there are downsides to the use of risk management: "In order to maximize net safety," the car "would position itself away from the large truck and closer to the small car, presumably because a crash with the small car would be less severe and safer overall" (Goodall, 2016, p. 817) further arguing that this transferral of risk without anyone's consent is unfair. Further to this, collisions will occur, there are situations where they are unavoidable for whatever reason. Any argument about not requiring an ethical system would absolve manufacturers from any liability concern, but would eliminate the possibility to minimise harm when a collision does occur, or act in a way that is deemed socially acceptable, hopefully providing society with more acceptance of the new autonomous technology.

Like risk management, research into Vehicle to Vehicle (V2V) and Vehicle to Infrastructure (V2I) communication is underway, to try to avoid collisions occurring at all. To briefly explain, devices around the road work together to identify potential risks. Due to these mechanisms being potentially further away from the vehicle at risk, there is more time to analyse and react. This alleviates the current issues with autonomous vehicles using close quarters detection mechanisms (such as radar and ultrasound) by being able to see issues from much greater distances (Knight, 2015). This was "successfully demonstrated" by Honda, with "the ability of a car equipped with Dedicated Short-Range Communications (DSRC) technology to de-

tect a pedestrian with a DSRC enabled smartphone" (Honda Government Relations, 2014).

Whilst V2V and V2I are good methods of fixing potential issues, existing road networks and vehicles do not have these technologies equipped, meaning little to no benefit to a vehicle that has V2V capabilities (Graham, 2017). This means that whilst there are still vehicles and infrastructure that do not have the capability, there will be black spots in the detection mechanisms, meaning the risk of collisions occurring is still probable, thus indicating vehicles do need some form of alternative mitigation in the interim.

6. Web Survey

The purpose of this study was to try and recreate a similar methodology to that of the MIT's Moral Machine, whilst adding independent variables of time and non-time pressure as well as actor. The goal was to identify if there is a significant difference between time and non-time pressure during collision scenarios, whilst also assessing if there is a significant difference between actor context on the decisions made. The benefit of isolating the data collection from the Moral Machine was to try to avoid differences in sample size and any unexpected discrepancies in methodology.

This chapter starts by describing the methodology used and subsequently considering the results that were gathered from the study. The discussion section is explained in a joint fashion with the second study due to both contributing to the final conclusion.

6.1 Implementation

This section described how the Web Survey and Random Scenario Generator were designed and developed, to meet the requirements of the methodology that is described later on in the document.

6.1.1 System and Software Design

In order for the web survey to be developed, requirements needed to be gathered. This was achieved by speaking to both supervisors and looking at the design style of the MIT's Moral Machine.

These requirements then allowed the progression to the design phase which involved identifying the technologies to use that would provide the best result, whilst also

yielding the best time-efficiency, ensuring the research continued at a reasonable pace.

To check the developments, the systems testing methodology which uses a "black box testing method used to evaluate the completed and integrated system" (Aebersold, 2019) was used. This tested both applications from start to finish to ensure they would meet the expected output from the design phase. Subsequently, this was further validated by acceptance testing, which was used to gain approval from stakeholders, ensuring the application was in line with the goals of the study. This was achieved by arranging a meeting with supervisors to test the web survey and provide any feedback they had on the implementation.

After this, usability testing was performed, which validated ease of use from the end-user's perspective. This was undertaken by asking around ten colleagues to test the application and identify if there were any issues and if they found anything difficult to understand. Testing participants undertook end-to-end testing which ran from what would be the start of the study to the end. This allowed the testing participants to understand the study as a whole and provide feedback.

Feedback gained gave the requirements for the evaluation phase, which would allow a new phase of software development.

Project Management

Development was kept on-track via the use of time management. In some projects, the benefits can be financial spending or other factors. For this project, money was not involved, the main cost was wasting time resulting in the studies being pushed behind.

Several milestones were drawn up, that would allow development to keep within the time-constraints.

Milestones to create a random scenario generator

- Creating SVG images for each character and asset

- Converting all images into PNG's
- Application can open PNG images
- Application can layer images over the top of one another
- Application can create images with all possible combinations of characters and assets
- Application won't recreate an image
- Application can save the images to Azure
- Run the application for 24 hours to generate huge quantity of images

Milestones to build the web application

- Development of the backend SQL Database to handle data storage of participants answers
- Development of the home page
  - Providing an ability to register onto the study
  - Providing anti-robot security by implementing Google ReCaptcha
- Development of the scenario selection page
  - Retrieve images from Microsoft Azure Blob Storage
  - Build suitable Model Binding to transfer data between client and server
  - Implement independent variables (time-pressure and actor)
  - Repeat application in static fifteen occurrence cycles
- Push the application to the Azure web app to test server reliability

Design Quality

A major issue with any form of software development, is design quality, one such area being 'Coupling', identifying how closely related individual pieces of data are to

one another. The ideal goal is that data is as weakly coupled as possible, essentially ensuring that data does not change unexpectedly when other data changes.

The web survey implemented weak coupling by ensuring that all images used were in their own file in Azure Blob storage, meaning files could be accurately retrieved and overwritten with no knock-on effect to other images.

Another design quality is 'Cohesion', the ability to measure how closely inter-related pieces of data are to one another (Easterbrook, 2001) which in itself assists with another design issue covered shortly known as 'Understand Ability'. Cohesion is the method of ensuring data is structured correctly, an example of which is Inheritance.

Cohesion was strong in the development of the web application. Data was kept in a strict parent-child relationship and data transmitted from client-side to server-side were model bound, ensuring that mapping data between the two environments was easy to visualise and understand. Understand ability was the most focused on design qualities in the application via the use of .NET Coding conventions (Microsoft, 2015).

Program Reliability and Efficiency

Program reliability is affected by a few key issues. Firstly, performance and scalability, this is where the application should be able to cope when under user load with as minimal application slow down as possible. This was not large concern for the web application because the application would come under less user load than an enterprise application.

Hosting the site in Azure meant that software could have as little downtime as possible to ensure users are not put off using the application. Azure web apps provide local redundancy meaning that should the server fail, the site would be automatically migrated to another server with no downtime.

The last area of program reliability that was considered was application fault. This is caused by poor code, user interfaces, application logic and page navigation issues, this leads itself towards poor user experience. This was the biggest program reliability consideration for both applications. This was handled by using a well-established,

open source style library called Bootstrap and built in error handling statements in C#.

6.1.2 Technical Information

Design Decisions

The following section explains the rationale for each of the components built into the project in order for the study to gain results and enforce the constraints of the independent variables. Where areas of code are fundamental to the constraints imposed, these will be included to assist, should replication of results be required.

Why a Web application was chosen?

A web application was chosen due to following similar design choices as the MIT Moral Machine. This was to limit as many factors that may cause unexpected effects on the study. The use of a web application provided the ability for the study to be conducted remotely, meaning participants could remain anonymous.

Vue

Vue is a JavaScript framework which provides the ability to design a responsive, single page, web application. Therefore, web pages only render parts of the page that have changed, rather than the entire HTML document. Vue, allowed the changing of the web page at run-time based on the independent variables of the study.

ASP.Net Core

ASP.Net Core, dramatically changes the code base of the .Net family. It provides the ability to host applications away from Windows Server environments supporting a cross platform model.

ASP.Net Core comes bundled with two core features, a backend programming language (usually VB.Net or C#) and a front-end programming language, Razor.

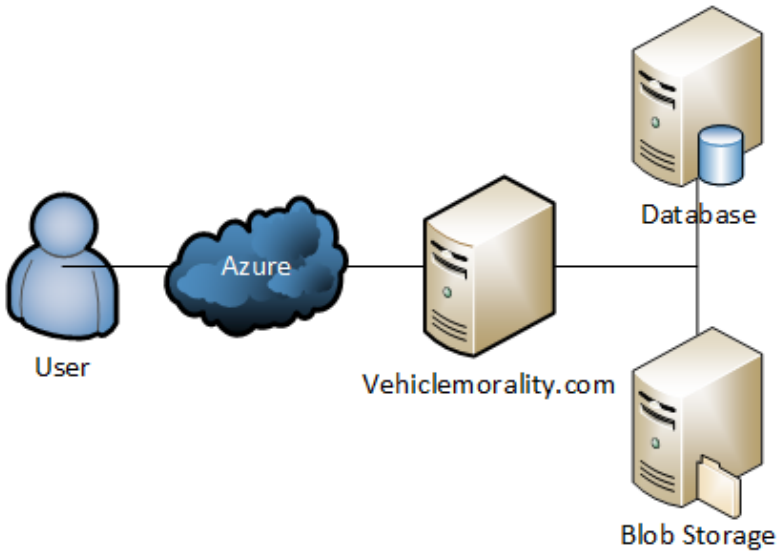

Figure 6.1: High level diagram of connected interfaces between user and azure resources.

6.1.3 System Structure

The application contains a few programming languages and techniques, designed to be utilised where best suited, or where simplicity is paramount.

Infrastructure Documentation

All hosting technologies used for this application were in Microsoft Azure. This was due to not having to understand networking and server management.

Within Azure the following technologies were used:

- Web App (the hosting platform for the developed application)
- SQL Database (the database platform which stored the results from the study)
- Azure Blob Storage (the storage account used to manage storing the images randomly generated for the application)

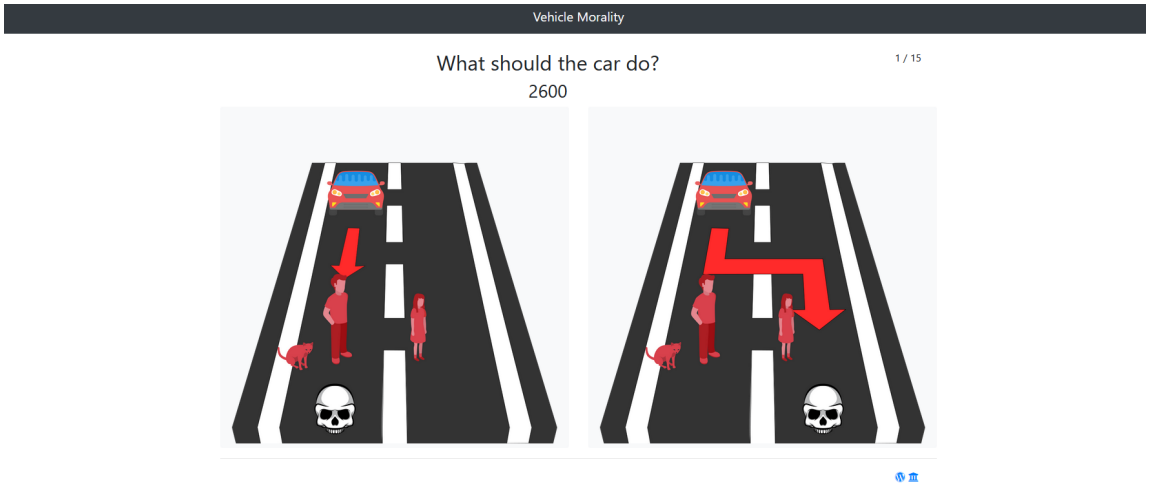

Figure 6.2: Example of the user interface shown to participants during a time constraint and autonomous vehicle actor scenario.

6.1.4 Random Scenario Generator - Technical Information

System Runtime Documentation

Unlike the web application, explaining system structure is not as necessary due to not utilising the MVC framework, or hosting within an environment. For this reason, more information will be provided around how the application created random generated images and uploaded to Azure Blob storage ready for use in the web application.

The application was written in a for loop which ran up to ten million times to produce as many image combinations as possible. Within this loop, an interrupt was installed and fired every thousand passes, to determine whether to continue creating images.

Before any image placing could occur, the system determined how many characters to place either side of the road via two random number generators which created the character count on the left and right side of the road. The system then selected characters for each side of the road, by randomising which characters to get.

From the characters, an image file name could be constructed creating a unique file name. This was achieved because the unique identifiers of the characters could be

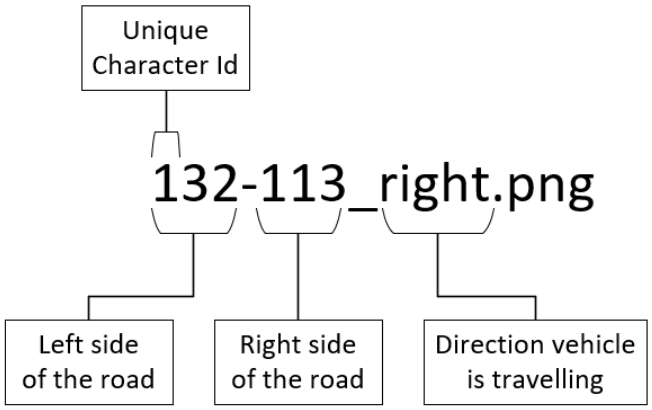

Figure 6.3: Example of the unique file names.

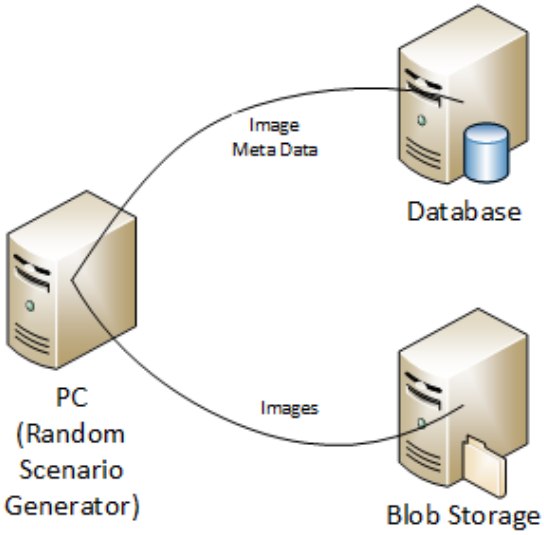

Figure 6.4: A high level diagram of how the random scenario generator sent data to connecting environments.

concatenated together in the order they were placed, and either side of the road, which produced meta-data about the image.

As well as the standard file names, the flipped image, bollard images, and illegal/legal image file names were generated.

- Flipped images, were where the characters on either side of the road, were reversed to the other side of the road.
- Bollard images were where characters on one side of the road were moved inside the on-coming vehicle and their prior position was replaced with a bollard. The

remaining characters were kept in their original position. Those images were also subsequently flipped to get both possible options.

- Legal/illegal images involved taking all images from the current pass, duplicating them, then inserting all possible combinations of legal/illegal effect on the image, in form of traffic lights, either red or green.
- Finally, all images had skulls placed in corresponding locations where death to the characters would occur from the on-coming vehicle.

For all of these images to be produced, the system had to overlay the characters onto the background, as well as having the direction arrows and vehicle being added.

To make scaling and positioning of the images simpler, the locations of each character were scaled and positioned prior to the images being overlaid. This meant that the images could be retrieved and placed, without any complex algorithms to scale and position, which could have introduced a multitude of bugs.

Pseudo Code

To assist in understanding the random scenario generator, pseudo code has been provided, highlighting the combinations that were used to build the scenes.

```
function buildCombination(Argument leftCharacters, Argument
    rightCharacters){
  For each character on left and right sides
    Get image based on position in road
    Add image to overall scene
  In copy of original scene, add all combinations of traffic lights
  In copy of original scene, add bollard to left of scene and move left
    characters into car
  In copy of left bollard image, add all combinations of traffic lights
  In copy of original image, add bollard to right of scene and move
    right characters into car
  In copy of right bollard image, add all combinations of traffic
    lights
  For each image
    Save to blob storage
```

```
    Add meta data to database
}

Do 10,000,000 times
  Get random number of characters on left of road
  Get random number of characters on right of road
  For each character on left and right sides
    Get random character type from database
    Set character number to position in road for meta data
  From meta data, check blob storage for pre-existing file
  If file does not exist
    Send the left and right characters to the buildCombination function
    Move left characters to right characters, and right characters to
    left characters.
    Send the flipped characters to the buildCombination function.
  Loop
```

6.2 Methodology

6.2.1 Primary Data

The intention of the web survey was to collect primary data specific to this research question because there is no data directly applicable and although the MIT's Moral Machine did cover a section of this research, it is not possible to infer the further environmental positioning and time-based pressures that may influence the results. This does come with its advantages, in that it is designed for the specific research question at hand (Hox and Boeije, 2005, p. 593).

6.2.2 Web Survey Design

To gain comparative data on each of the different factors being tested in the hypothesis, participants were randomly assigned to a factor set that would determine how the study behaved in front of the participant. For non-time pressured scen-

arios, participants were presented unlimited time to complete each scenario. In contrast, participants under a time-pressured situation were given five seconds to ready themselves, with the secondary intention to give their web browser time to ready everything on the screen. Participants then had three seconds to decide, double that of average time-to-collision findings where "the average TTC that braking was initiated at was found to vary in the sample population" of 47 "from 1.1 to 1.4 seconds" (Kusano and Gabler, 2011, p. 435). The reason for this increase in decision time was the further expectation of web browser issues. Therefore, to reduce risk of unnecessary indecision, the value was increased. Should participants fail to complete the decision in time, their result was registered as selecting the left image and the vehicle continued down the road, hitting whatever was in-front of it.

	Time	Non-Time
Bystander	Group 1	Group 2
Self-Driving Car	Group 3	Group 4
Driving	Group 5	Group 6

Table 6.1: Different groups participants could be assigned too.

To meet the requirement of assessing if there is significance between environmental positioning, the participants were informed prior to the study, and on each scenario.

The survey requested participants to complete fifteen, randomly generated scenarios that were presented to them iteratively. Each scenario consisted of two images, providing a pair-wise comparison between the two, which represented the two outcomes a participant could choose. One option was a situation where the car continues down the path in which it was already progressing. The other was where the vehicle would swerve into the other lane. Participants were shown the potential outcomes of each of the decisions via symbolic skulls over the character's heads within the images.

Each side of the road consisted of a random number of people between the range of one and five, which could consist of any of the seven-character types in a random configuration:

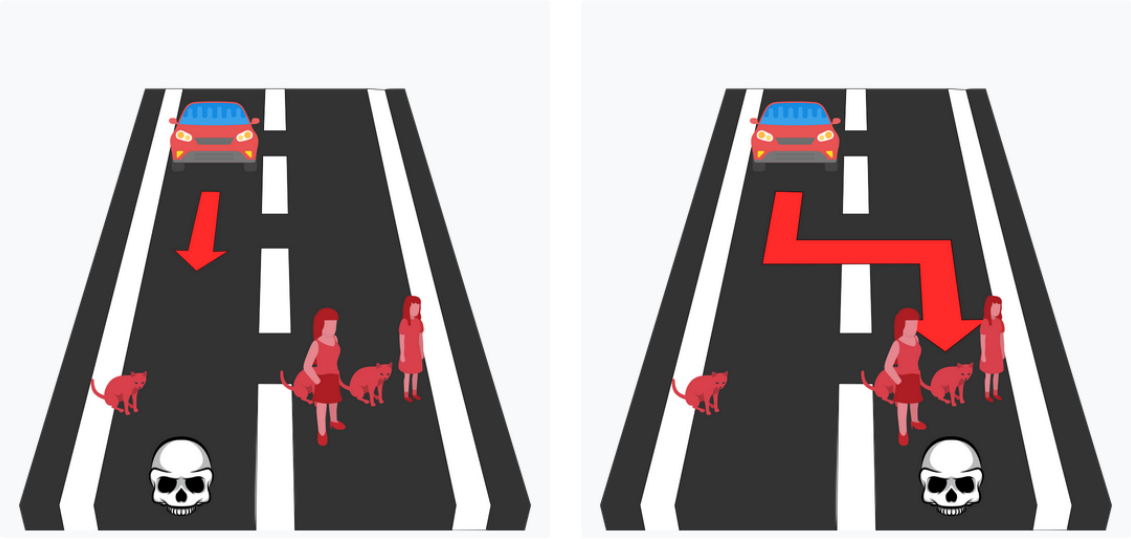

Figure 6.5: Example of a left to right image presented to participants.

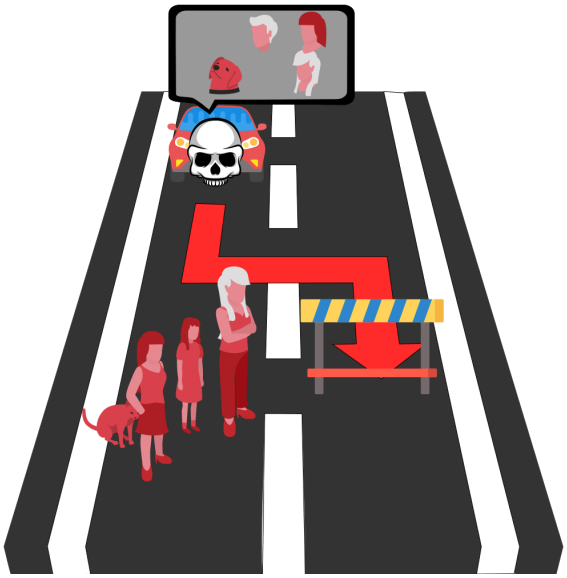

Figure 6.6: Example of a right hand choice collision when a bollard is presented to participants on the right side of the lane.

- Adult Male
- Adult Female
- Elderly Male
- Elderly Female
- Male Child
- Female Child

- Dog
- Cat

It was possible for one of the sides of the road to be overridden with a bollard instead of characters, with the characters on the right side of the road being moved into the vehicle. It is also important to mention that the remaining left-hand characters were flipped onto the right-hand side to produce images with the bollard on the left and the right. This meant participants could also be presented with situations where they had to choose between self-sacrifice and self-preservation. Images could also contain traffic lights on each side of the road, which could consist of being green, or red; images did not contain scenarios where traffic lights were only present on one side of the road.

Due to a bug in the study, adult male characters were not added to the study, meaning a bias in results was present. To alleviate this, the study was re-run with adult females removed and adult males added in. Via the use of generalised estimating equation, it was then possible to see people's preferences of those that are more likely to be saved or killed, whilst ignoring the prior bug in the system.

6.2.3 Sample Design

The sample was very open for this study, allowing anyone consenting they are over the age of 16 and with no severe mental health issues (e.g. PTSD) to participate and with the limitation that only those with internet access were able to complete the study. This reduces the potential sample size. It is also important to note that as of 2016, internet access across the globe is still largely used by higher wealth countries (Poushter, 2016). Therefore it can be safely assumed that the loss of participants who would be impacted within a number of years, would be far less than initially expected. A limitation of web surveys with regards to mental health disorders is that it is not possible to prove someone does or does not have an issue. Instead, the suggestion must be assumed to be advisory; therefore the mental health statement could have been ignored, but there is no evidence to prove so.

Participants were kept completely anonymous through a multitude of factors. Firstly, the participants never had direct contact with the researcher. Secondly, the data gathered contained no way of tying the information back to one person. This was because the only "personal" data being gathered was demographic information which participants provided themselves. They had every option to falsify or not provide this information, should they wish to. The choice of anonymity has the advantage of attracting participants to select what could be argued as a sensitive decision. This could be classed as sensitive because they would be selecting between the deaths (although virtual) of individuals. The difficulty is that the responses given by the participants can never be followed up. This means that should the information provide some level of interest, the gathering of more in-depth information would need to be left to a separate study (Vaughn, 2017).

6.2.4 Participant Recruitment

To gain participants for the study, a number of messages to social media sites such as Facebook and Twitter were sent. The Facebook messages were also shared by supervisors to attract more attention. Further to this, a blog post on Tumblr, and a Wordpress blog site were created to try to attract further attention. Based on the sharing features of Wordpress, these blog posts were also shared onto Facebook and Twitter at the time of release. The next method of participant recruitment involved posting onto multiple sub-reddits specifically tailored or requesting research and survey participants. The last method was using the University's Staff News portal to request participants for the study. Due to all of these being an online format, and being anonymous, it is not possible to tell the click rate for these and how many participants took part via each method.

6.2.5 Study Procedure

Whilst information about the allocation of the independent variables is provided earlier in the document, it is necessary to describe the entire procedure participants

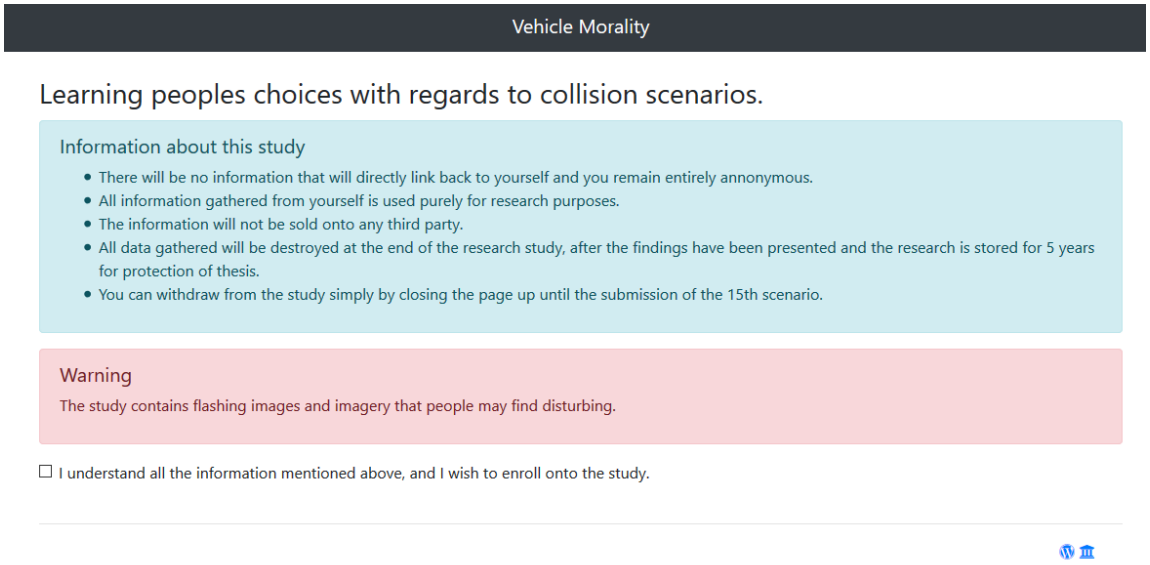

Figure 6.7: Home page shown to participants before they choose whether to undertake the study.

took when undertaking the study. Participants would first access the website. This would present them with initial information about the study. Should participants wish to take part, they would tick the checkbox at the bottom of the page to indicate they understand the information. This is shown in figure 6.7.

Participants would then fill out their demographic information, hitting submit to begin the study.

Participants would then be shown more detailed instructions of what their task involved, for example what actor they were taking part as. Participants had as long as they liked to read the information, present the button to indicate they are ready. an example of this is shown in figure 6.8. This started the fifteen scenarios for participants to select.

6.2.6 Study Analysis

From the data collected in the study, the intention was to identify if there were significant differences in participant decisions via the use of generalised estimating equations. Identifying the level of diversity between demographic decisions (Karpov, 2017, p. 754).

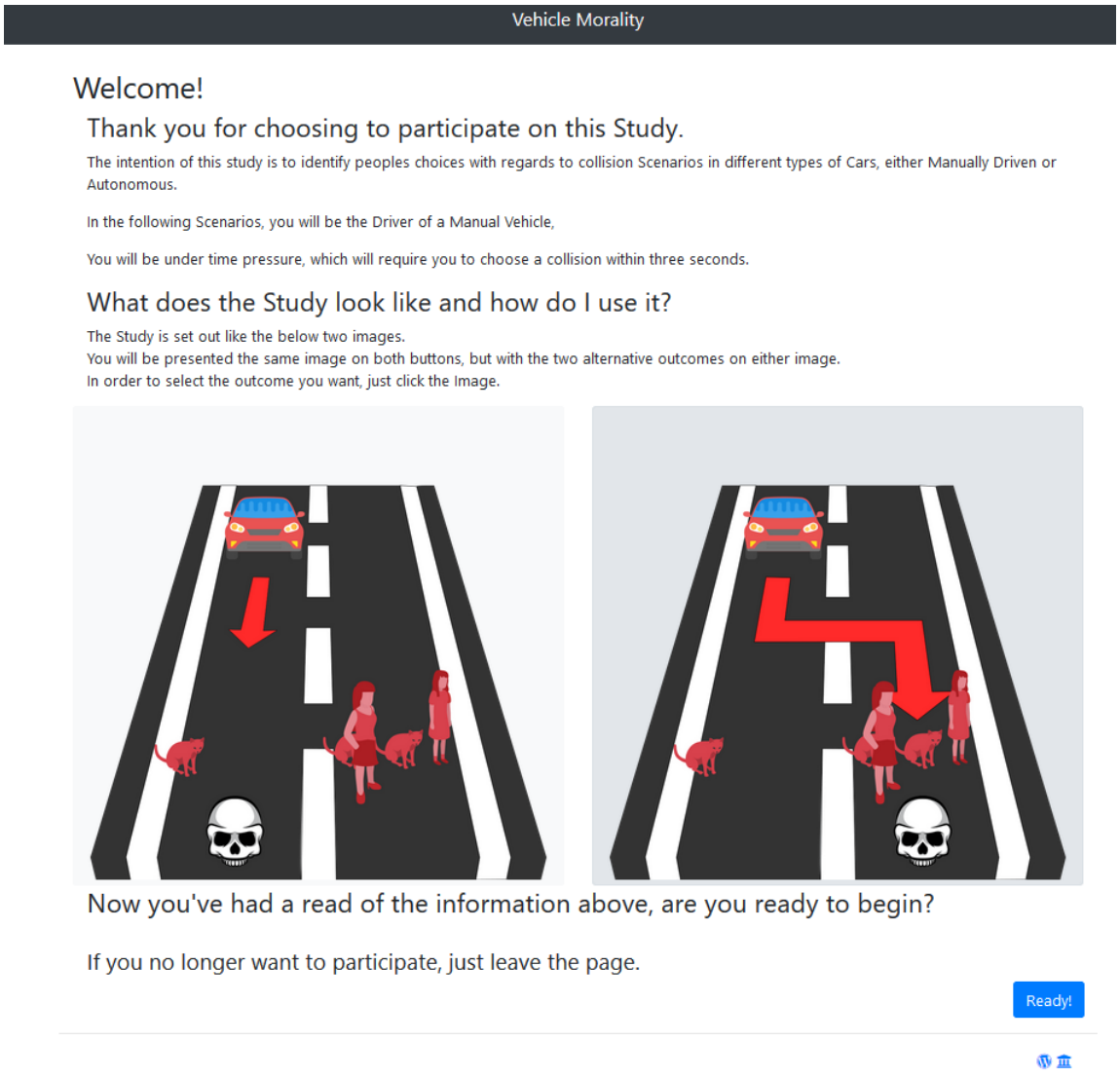

Figure 6.8: Information presented to participants before they begin the fifteen scenarios.

Generalised Estimating Equation (GEE) is a statistical analysis method that "estimates population-averaged"(Hong and Ottoboni, 2017a) polling data. Polling data can vary on a number of factors. For example, population density and variations in options to select. GEE is able to statistically ignore the variations and provided an estimated value of deviation from a normal value also known as an intercept. In results, the intercept value is indicated by a B. The more positive the B value, the more in favour of an evaluated characteristic. Thus, the more negative the value, the further against the characteristic. These B values can then be identified as random values, or significant based on the P value given off by GEE. Should the P values be indicated as significant, it is possible to infer that the characteristic causes a deviation to the intercept value.

To further build an understanding of what participants chose, a progressive addition of interactions were carried out to analyse see if one factor has an impact on the rest of the data. This constituted the independent variables, with the use of demographics to examine the effect characters had on participant preferences. The benefit of GEE and "having panel data (repeated measurements) like this is that we can control for time-invariant, unobservable differences between individuals. Having multiple observations per individual allows us to base estimates on the variation within individuals" (Hong and Ottoboni, 2017b) thus providing the ability to see what impacts a person's decision.

The data was captured based on the position of each character on the road. If the character was on the left of the road, they were recorded with a -1, with the right-side being a 1. Comparatively, participants decisions were recorded with the same values to indicate which side of the road they selected. Should a participant select the same direction of the road that a character's position was on, GEE will change the outcome by a positive value, indicating someone is more likely to go for that demographic type. Subsequently selecting the opposite side of the road will have a negative effect on the character due to it being regarded as a potential effect for causing participants to select the other option.

6.3 Results

To analyse the data gathered in the first study, it needed to be split into two versions. One analysis for images without bollards, or "standard images" and the other analysis for bollard images. All of these utilise Generalised Estimating Equation.

Overall 202 participants undertook the study with fifteen scenarios completed. Those that did not complete fifteen scenarios were ignored due to assumption they withdrew from the study. All participants consented they were over the age of 16.

6.3.1 Standard Images

Ignoring Independent Variables

To gain an understanding of the data, comparison of characters and lane selection was undertaken without independent variables being considered. This enabled the ability to see the overall participant decisions. Via the use of the generalised estimating equation, it is possible to see that regardless of independent variable, participants are more likely to respond in a utilitarian manner, saving the more over the few. This is visible by looking at the intercept of 6.2 ($B = -0.484$, $p = 0.008$) which indicates participants are more likely to swerve from people in the road, than hit them. This lines up with the results from the Moral Machine which also showed "stronger preferences" for "sparing more lives" (Awad et al., 2018, p. 60), and begins to verify that the Moral Machine's results are valid to general human viewpoints.

Further to this, the generalised estimating equation provides information as to which characters were more likely to be targeted or killed. In the case of ignoring independent variables, the cat came out as causing less of an effect in changing a participant's decision ($B = -0.488$, $p = 0.004$), whilst still showing that a cat will cause a person to swerve. Human characters, in contrast, cause a heavy weighting on changing a participant's decision, with young male characters causing the biggest effect on participants decisions ($B = -1.693$, $p < 0.001$). Out of the human characters, old

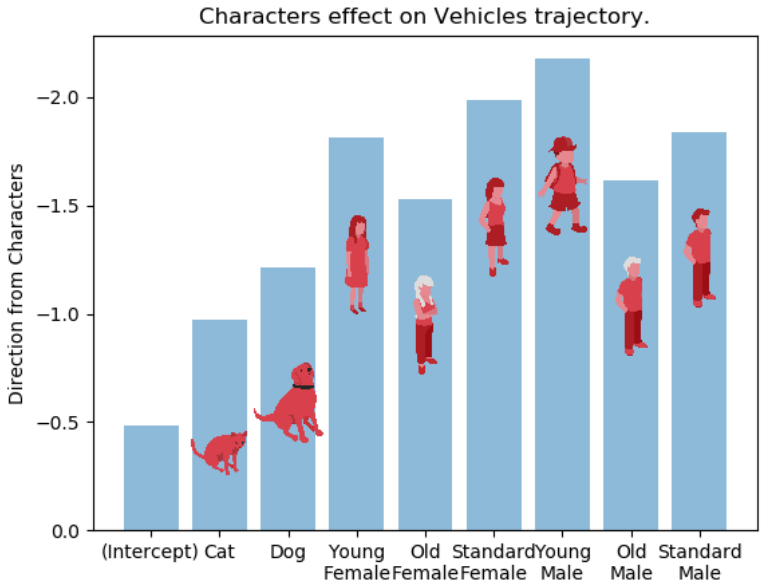

Figure 6.9: Chart showing the effect characters have on a vehicles trajectory.

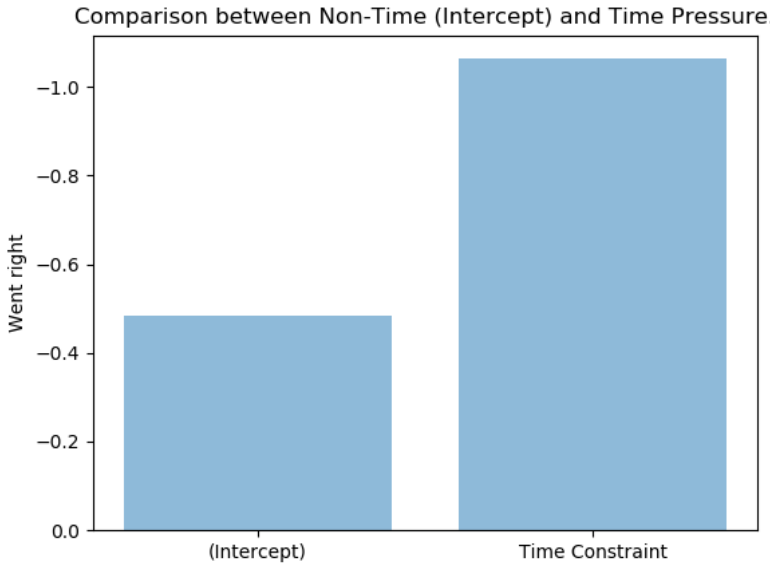

Figure 6.10: Chart showing the effect time had on likelihood to intervene.

Parameter	B	Std. Error	Hypothesis Test Wald Chi-Square	df	Sig.
(Intercept)	-.484	.1836	6.947	1	.008
Cat	-.488	.1678	8.446	1	.004
Dog	-.731	.1571	21.684	1	.000
Young Female	-1.329	.2601	26.100	1	.000
Old Female	-1.046	.2224	22.113	1	.000
Standard Female	-1.501	.3063	24.001	1	.000
Young Male	-1.693	.2847	35.353	1	.000
Old Male	-1.134	.2171	27.293	1	.000
Standard Male	-1.356	.2671	25.754	1	.000

Table 6.2: Generalised estimating equation results looking at character effect on a participants decision without time pressure.

Parameter		B	Std. Error	Hypothesis Test Wald Chi-Square	df	Sig.
(Intercept)		-.484	.1836	6.947	1	.008
Time Constraint		-.579	.1912	9.183	1	.002
Time	Cat	-.085	.1571	.292	1	.589
	Dog	-.187	.1878	.994	1	.319
	Young Female	.373	.2802	1.774	1	.183
	Old Female	-.067	.2446	.076	1	.783
	Standard Female	.024	.2907	.007	1	.934
	Young Male	.343	.3021	1.288	1	.257
	Old Male	.220	.2374	.859	1	.354
	Standard Male	-.116	.3264	.125	1	.723

Table 6.3: Generalised estimating equation results looking at character effect on a participants decision under Time Pressures.

characters cause the least effect on making people swerve. This would indicate that in a collision scenario between young characters and old characters, there is a higher chance that someone would sacrifice the old people for the young people.

When looking at the effect of traffic lights on the environment, it is clear this has no impact on participants decisions due to results showing participants are more likely to head towards a green light, but with the results being non-significant ($B = 0.019$, $p = 0.927$); it is only logical to class this indication as being random.

Time Constraint

Moving onto comparing the time pressure independent variable, it is possible to see that those under non-time pressure are more likely to intervene in a collision than time-pressured participants. This is to be expected due to time-pressured participants being able to timeout, causing their decision to default to continuation. This is a shortcoming of these results because the data does not account for decisions where a participant timed out. However, it is possible to argue that the strong significance of this result shows there to be some interest in the outcome.

When comparing if time-pressure caused a difference on which characters are hit, the results are found to be non-significant. This further shows time-pressure as being a factor which is not important in the Trolley Problem's ethical dilemma. However, when comparing the effect time pressure has on young characters, there is a sign that those under time pressure, are more likely to hit them. On the flip side, old characters are more likely to be hit when under non-time pressure and have a higher chance of being saved in time pressured scenarios. Both of these results are non-significant and so can only be assumed as random outcomes.

6.3.2 Bollard Images

Unlike the standard images, the dependent variable was based on whether participants chose to go for the bollard.

Ignoring Independent Variables

When looking at results shown in figure 6.4, there is a clear significance that participants are more likely to go towards the bollard than away from it ($B = 0.526$, $p = 0.014$).

Further to this, changing the side of the road that the bollard is on, shows no significant effect on the choice participants make ($B = -0.005$, $p = 0.971$).

Parameter	B	Std. Error	Hypothesis Test		
			Wald Chi-Square	df	Sig.
(Intercept)	.526	.2131	6.092	1	.014
Bollard Position	-.005	.1356	.001	1	.971
Cat	.265	.1654	2.565	1	.109
Dog	.046	.1408	.108	1	.743
Young Female	.234	.1383	2.866	1	.090
Old Female	.268	.1403	3.655	1	.056
Standard Female	.593	.1530	15.045	1	.000
Young Male	.437	.1428	9.378	1	.002
Old Male	.536	.1471	13.267	1	.000
Standard Male	.461	.3114	2.187	1	.139

Table 6.4: Generalised estimating equation results looking at participants likelihood of hitting the bollard.

Without any independent variables, there is a mixed significance between the character types. Cats, dogs, female old, female young and male standard are all classed as non-significant but do cause a weighting towards the bollard.

However, female standard ($B = 0.0593$, $p =< 0.001$), male young ($B = 0.437$, $p = 0.002$) and male old ($B = 0.536$, $p < 0.001$) have a significant effect on participants decisions causing them to be more likely to go towards the bollard.

It is also important to note that the B values for the character types are visible as positive. This indicates that participants still respond in a utilitarian manner when there is the assumed view that life both in the vehicle and outside of the vehicle is equal. This is due to both characters inside the vehicle and on the road, have a weighting in the table.

Time Constraint

When comparing the effect of time constraint on the participants decisions, there is a significant effect that time pressure does cause participants to go towards the bollard more, regardless of which side of the road they are on ($B = 0.512$, $p = 0.012$).

When assessing the interaction that time pressure has on participants decisions to avoid the characters, there is no significance across the character types. This is with the exception of female standard which shows a marginal significance of causing

Parameter		B	Std. Error	Hypothesis Test Wald Chi-Square	df	Sig.
	(Intercept)	.526	.2131	6.092	1	.014
Time	Old Female	-.120	.1524	.615	1	.433
	Standard Female	-.364	.1722	4.464	1	.035
Non-Time	Old Female	.268	.1403	3.655	1	.056
	Standard Female	.593	.1530	15.045	1	.000

Table 6.5: Generalised estimating equation results looking at participants likelihood of hitting the bollard when under the effect of time pressure and female standard characters in the environment.

participants to swerve away from the bollard more, see figure 6.5. This could be either due to females being in the vehicle, or less females on the road compared to other factors in the vehicle. ($B = -0.364$, $p = 0.035$).

Actor Constraint

When considering the effect perspective has on participants decisions, it is clear there is no significance on their likelihood of changing decision.

6.4 Conclusion

This chapter described and explained the rationale for the methodology used in the web survey by inducing time constraints of three seconds on participants decisions. The methodology further explains the inducement of the actor independent variable.

The methodology then considered the results of the chapter. This showed that participants are more likely to react in a utilitarian manner, regardless of characters acting as environmental stimuli. Further to this, the addition of time constraint and actor positioning has no significant impact on participants decisions.

This chapter also highlighted that when participants are presented with a self-sacrifice and self-preservation comparative scenario, participants are likely to sacrifice themselves. This continues to be the case when under the effect of the two independent variables.

7. Driver Decisions in a Simulated Environment

After the identification of the non-significance between both independent variables, the study was focused on areas that indicated the possibility of causing a difference on decisions. This study employs the use of virtual reality to provide a more immersive environment, to create a more realistic situation that participants may adapt their answers to.

The study compares time and non-time pressure independent variables against participants decisions to an equal death ratio of 1 to 1 when against the choice of self-sacrifice or self-preservation. Firstly, this chapter describes the methodology used to gain the results later identified in the chapter. The results section then delves into quantitative, generalised estimating equation and cross-tabulation. It further investigates via thematic analysis on participants responses to questions around the decisions made. This provided more qualitative insight into the decisions.

7.1 Implementation

This section described the implementation of the virtual reality environment, highlighting implementation rationale.

7.1.1 System and Software Design

Like the web survey, requirements were gathered by speaking to supervisors and analysing the design of the Moral Machine.

These requirements then allowed the progression to the design phase which involved

identifying the technologies to use that would provide the best result, whilst also yielding the best time-efficiency, ensuring the research continued at a reasonable pace.

Similar to the web survey a "black box testing method used to evaluate the completed and integrated system" (Aebersold, 2019) was used. This tested the application from start to finish to ensure they would meet the expected output from the design phase.

Subsequently, this was further validated by acceptance testing, which was used to gain approval from stakeholders, ensuring the application was in line with the goals of the study. This was achieved by sending recordings of the VR environment to supervisors to provide input on the development.

In tandem with this, usability testing was performed, which evaluated any issues that users may come across when using the system. This was conducted by asking five colleagues to test the application from start to finish. This identified that because the use of keyboard controls: left and right arrows to move and space bar to start, made the application difficult to use because people in VR could not see where their hands were. The next iteration split the keyboard in half and allowed testers to hit whichever side of the keyboard they wanted the car to move. Issues around the shift and alt keys caused unexpected changes to the control scheme, meaning it failed frequently. The final version used an Xbox One controller. Testers were able to keep their hands in the same place and this fixed the prior issue.

Project Management

Development was kept on-track via the use of time management. In some projects, the benefits can be financial spending or other factors. For this project, money was not involved, the main cost was wasting time resulting in the studies being pushed behind.

Milestones for the development of the simulated environment

- Placing road assets in environment

- Placing buildings and objects in environment
- Placing character in environment, with suitable animation
- Develop scripts to be able to handle user selection
- Provide visual cues around user selection constraints
- Develop script to handle global variables to track within group design status

Design Quality

A major issue with any form of software development, is design quality, one such area being 'Coupling', identifying how closely related individual pieces of data are to one another. The VR environment utilised weak coupling by ensuring data values were passed by value, not by reference, meaning a new version of the variable was manipulated.

Another design quality is 'Cohesion', the ability to measure how closely inter-related pieces of data are to one another (Easterbrook, 2001) which in itself assists with another design issue covered shortly known as 'Understand Ability'.

The VR simulation contained data transactions between scripts which were inter-related, meaning each script became reliant on the other. Although this did not cause issues to the run-time of the application, the next design quality, Understand Ability, became forfeit due to the fact Unity requires scripts to be added as components within a game asset to function. This meant remembering which game object contained the script became difficult over-time. To rectify this, scripts were kept inside a single game-object to make finding scripts easier.

Program Reliability and Efficiency

With game development, especially virtual reality, efficiency and performance is key. Failure to ensure the framerate of the application can remain high enough, can potentially cause motion sickness in VR (Suarez, 2018). To reduce the risk of frame

Figure 7.1: View in VR of the game environment when reaching the collision.

drops, assets were only used in the environment where necessary, saving processing power.

7.1.2 Technical Information

Design Decisions

Why a Unity Game Environment?

There was no prior knowledge of game development and due to prior experience with C# decreasing the learning curve, Unity was a suitable option because of its out-of-the-box VR support, meaning less need to understand the complexities of VR tracking.

Game Object Layout

The application's core components were separated into their own parent game objects. These consisted of:

- Road – The game object dedicated to handling road objects.

Figure 7.2: View in VR of the game when selecting to collide on the right-hand side.

- BlueSuitFree01 – The male character used for collision scenarios.
- Car (Main Character) – The vehicle that users were sat in during the scenarios.
- Directional Light – The Unity light object used to give shadows to the environment.
- Buildings – The parent game object handling all buildings placed in the world.
- Barrier – The parent game object handling all of the barriers shown in a collision scenario.
- Terrains – The parent game object handling the two terrain colliders.
- Spline – The parent game object storing all the CurvySpline game object to create the on-rail effect of the vehicle.

Script Assets

All Scripts except the Curvy Spline Controller were stored in the car game object. As explained in the Design Decisions section, this was to aid maintenance. Each of the controllers is explained below.

Car Controller

This was the controller provided by the Unity Standard Assets Package on the Unity Store. This script was repurposed for its integration with the audio system for the car, as well as its pre-configured rev algorithm, meaning the audio and gear changes could be handled with little programming. The script was repurposed to house the vehicles speed and rpm needles which would provide a more realistic look to the vehicle.

File Handler

This script was developed to handle variables between scenes. Json.Net was used to store the global variables and read them in each time the scene was loaded. This also meant recording the order of scenarios after the study was achieved.

Input System

The input system controller handled the outside game environment, rather than the vehicle itself. This controller configured the independent variables of the study as well as changing the position of the character and barriers based on the Json.Net files results.

Heading Calculation

This script was designed around the display panel used inside the vehicle to display a countdown until the collision would occur. As well as this, the slow-motion activation was stored and run in this script.

Decision Engine

Decision Engine was used by the CurvySpline system to be able to change the direction of the rails when a user made their decision. As well as this, the controller would lock down if the user had made their decision which would stop them from flicking the car around the different rails.

Custom Audio

This script required the car controller in order to run, also part of the Unity Standard Asset Pack controlled the audio that the car made via the use of four audio clips which were combined to create a smoother audio experience when accelerating and decelerating.

Curvy Accel

This script was designed to handle the speed of the car, which was subsequently updated in the car controller for the audio effect. This script increased the speed of the vehicle by 0.9 Miles Per Hour (MPH) until the vehicle reached a trigger speed of 15 MPH, at which point causing a faster speed increase of 1.1 MPH. Once the vehicle reached 40 MPH, the vehicle stopped accelerating and maintained the speed up until the collision occurred. At the point of collision, the screen would then black out for participants to ensure they did not witness the outcome.

Spline Controller

This was an automatically generated script by the Curvy Spline package which handles the logic of the on-rails effect.

Slow Mo

When triggered by an external controller, this script would initiate the Slow Mo sequence that would enable participants to make a collision choice when under non-time pressure constraints.

7.2 Methodology

7.2.1 Primary Data

The intention of the virtual reality (VR) environment was to collect primary data to expand the results gained in the prior study. The reason for this, was to the researcher's best knowledge, there was no data directly applicable to what was gathered in the prior study. As with the prior study, this has the advantage of being tailored towards the research question (Hox and Boeije, 2005, p. 593). Another reason was that the prior study had no ability to follow-up from the data collection, meaning no qualitative understanding was achievable. The use of the VR study allows the gathering of this qualitative data to better understand participants decisions.

7.2.2 Virtual Reality Design

From the previous study, discussions were held to consider the results and next steps. One of the main areas of discussion, was the shortcomings of the first study. The main issue was the design of images. This was raised because the actor independent variable was only explained to participants in textual format rather than visual, unlike that of the time-pressure independent variable, and this could explain why the actor independent variable was found to be non-significant. The decision was to focus the following study on the independent and dependent variables that suggested areas of further investigation:

- Time-Pressure Vs. Non-Time Pressure
- Self-Preservation Vs. Self-Sacrifice
- Intervention Vs. Continuation

From this, the hypothesis from the prior study was carried over, with the actor independent variable removed, and immersion being added as a constant.

7.2.3 Study and Procedure Design

For the virtual reality environment to be used, the study was run on a computer in the University of Lincoln's ICT Department's meeting room. The University of Lincoln offered the potential benefit of being able to target a large sample set of at least a thousand staff members, with a high quantity of students being added into that sample set. This meant that results were solely of British citizens and had no comparison from other nationalities.

Due to financial limitations, participants were not offered payment for their participation, reducing the sample size down due to participants having to give up their own time to complete the study. Further to this, the time of the year that the study ran, coincided with students taking exams or leaving for the summer period; this meant a large sample size had left Lincoln, making it harder to find participants.

Participants had to be over 18 years of age and be free from mental health issues such as Post Traumatic Stress Disorder (PTSD) or physical impairments, such as blindness. This was to ensure participants were not negatively affected by the study, thus ensuring the study was ethical. Participants were not required to have a driving license or be able to drive a car. This was due to the study having an autonomous vehicle, so participants only had to press two buttons, and did not have to understand how to drive a real car.

To gain participants, a message on multiple social media sites (Facebook, Twitter and LinkedIn) were posted to increase awareness of the study occurring. Participants were also requested via an inter-departmental email which had an audience amount of around sixty potential participants, as well as an all staff research request via the staff news blog site. Further to this, was the ability to discuss the study with other members of staff, which yielded further potential participants.

Time to conduct the study was discussed and arranged with each participant which was anticipated to last around half an hour. Splitting the study into five minutes for the VR phase and the rest of the time devoted to interviews to gain qualitative data.

During the study, the purpose of the study, participant rights and the controls were explained to participants. This was because participants in VR were unable to see the controls, it was deemed as necessary to avoid confusion. Participants then consented they were ready to begin, by pressing the "A" button on the Xbox controller which would start the car on its journey. To then allow participants to get used to the motion of the car, the car pulled out of a parking space at a slow speed and linearly increased in speed every frame by 0.9 miles per hour (mph). The car increased to around 40 mph, before capping the speed limit. This was beneficial because limiting and increasing the speed meant factors were the same for all participants. This was the principle reason for using on-rails mechanics, instead of using Unity wheel colliders.

Participants would then make four decisions with a character being present on one side of the road, and a bollard on the other. The four scenarios allowed for the reversal of the character and bollard to subsequent sides of the road and the variation of the time-pressure independent variable. In order for participants to decide, they had to press one of two bumpers, or shoulder buttons known as "Lb" and "Rb". Pressing "Lb" kept the car on the left side of the road, thus pressing "Rb" moved the car into the right lane for the collision.

After all of the collisions had occurred, participants were requested to remove the virtual reality headset and were given time to ready themselves for the audio interview, should they wish. After consent from participants to begin the audio recording, participants were asked several questions via a semi-structured interview script. This enabled all questions to be asked, but allowed adventure into other areas, and more detailed explanations to be requested where necessary. Due to the lower sample size that came forward for the study, the use of interviews was beneficial in the "exploration of the perceptions and opinions of respondents regarding complex and sometimes sensitive issues and enable probing for more information and clarification of answers." One of the potential issues around semi-structured interviews is the risk of variation in words used to ask and elaborate on questions to participants, potentially causing a variation in responses. Semi-structured interviews make clear that

the effect is not on the words used, but the meaning that is conveyed from them which standardises the responses given (Barribal and While, 1994, p. 330).

7.2.4 Sample Design

As with the first study, the allowed sample size was relatively broad, with the minimum age being eighteen and anyone without physical or mental impairments such as PTSD being able to participate. Unlike the web survey, participant recruitment was centred on the Lincolnshire area due to the difficulty with transportation of the virtual reality equipment and time constraints on the study.

Participants were kept anonymous via numeric identification with participants having the ability to withdraw from the study at any point via the use of a secret word they provided, which gives the results a level of protection to avoid other people from removing other participants results. To further maintain anonymity, participants were never addressed by name during the audio interview. Their name was only recorded in the ethical consent form due to University of Lincoln ethical guidelines.

7.2.5 Study Analysis

Due to this study being both qualitative and quantitative, two methods of data analysis were required which would allow the individual analysis of both forms of data collection.

Qualitative Analysis - Thematic

When analysing qualitative data, audio recordings were transcribed to allow for thematic coding to occur. Initially notes were taken from a small sample size of the transcriptions to identify common themes that were present in the text. These were split into two parent categories called “viewpoints”, (quotes around participants general views on autonomous vehicles) and “outcomes” (the quotes around the decisions participants made).

To ensure the themes identified were reliable, an independent PhD student was requested to identify key themes of the transcriptions, without being prior informed of the themes identified at the time. On receiving the themes, the student identified an extra theme which was not initially identified. However both of the other two themes were validated. This produced the following three key themes:

- Viewpoints about autonomous cars
- Moral decisions during the study
- The application of the decisions to real-life collisions

When considering the initial common themes for moral decisions during the study, the following were easily identifiable:

- Learning the study
- Looking for more options
- Mistakes
- Regret
- Safety features
- Unwillingness to hurt a person
- Willingness to hurt a person

When considering the initial common themes for viewpoints about autonomous vehicles, the following were found:

- Blaming human driving
- Diverting attention from driving
- Excitement of new technology
- Losing ability to drive
- More testing
- Not being in control

Finally, when considering the application of the decisions to real-life collisions, the following were found:

- Changing decision compared to gameplay
- Keeping decision similar to gameplay
- Unsure on the decision

Each transcription was then coded into these nodes which would then allow for focused analysis to occur, as documented in the results section.

Quantitative Analysis

The purpose of the data was to understand whether time-pressure had an impact on whether a participant chose the person. To identify this, the data was analysed using Generalised Estimating Equation (GEE), as used within the prior study. This was done because the time-pressure scenarios had the ability to time-out. It provided the ability to remove these from the results, which GEE compensated against by understanding the probability the participant would make a similar choice.

The data was captured based on the position of the character on the road. If the character was on the left, then a value of -1 was used. When the character was on the right, the value of 1 was used. Subsequently, the data for a participant's decision was recorded with a Boolean false for left, and a true for right. Should a participant select the same direction of the road that a character's position was on, GEE would change the outcome by a positive value, indicating someone is more likely to go for the character. Conversely, selecting the opposite side of the road will have a negative effect on the character due to it being seen as a potential effect for causing participants to select sacrificing themselves.

7.3 Results

7.3.1 Quantitative

Forty participants took part in the study. All participants were British Citizens over the age of 18 and all were living around the area of Lincolnshire. This provided one-hundred and sixty responses to the collision scenarios.

The data that was analysed was used to identify if there is a significant difference between time and non-time pressure scenarios between self-sacrifice and self-preservation-based scenarios. To gain an initial understanding of the data, a cross-tabulation was run to identify any clear areas of choice difference between the independent variables.

Was Time Pressure			Hit Person: Missed	Hit Person: Hit	Total
Non-Time	Character Position	Left	82.5%	17.5%	100.0%
		Right	60.0%	40.0%	100.0%
	Total		71.3%	28.8%	100.0%
Time	Character Position	Left	67.5%	32.5%	100.0%
		Right	65.0%	35.0%	100.0%
	Total		66.3%	33.8%	100.0%
Total	Character Position	Left	75.0%	25.0%	100.0%
		Right	62.5%	37.5%	100.0%
	Total		68.8%	31.3%	100.0%

Table 7.1: Crosstab Results from Study Two.

In 7.1 there is no strong identification of difference between time and non-time pressure due to both having similar percentage values. What is unexpected is when participants are not under time-pressure and the character is on the right; there is more of a chance that the participant would hit the person than when under time-pressure. Initially this could be caused from a data issue, however the data was validated by running the same cross-tabulation on the raw JSON data before being converted to an SPSS file.

When identifying significance, the Generalised Estimating Equation was used to assess participant preferences between the independent variables.

When looking at the intercept value, non-time pressure, of 7.2 shows a clear significance, with participants being more likely to avoid hitting the person ($B = -0.978$, $p = 0.001$). When comparing the likelihood of hitting the person when under time pressure, the value increases, such that participants become slightly more likely to hit the person, however still likely to avoid the person in most cases. However, this value is insignificant ($B = 0.303$, $p = 0.293$). Regardless of whether this value is significant or insignificant, the results still indicate participants are more likely to sacrifice themselves than harm another individual.

			Hypothesis Test		
Parameter	B	Std. Error	Wald Chi-Square	df	Sig.
(Intercept)	-.978	.2831	11.934	1	.001
Character Position	.573	.2419	5.603	1	.018
Was Time Pressure	.303	.2884	1.104	1	.293
Character Position X Was Time Pressure	-.517	.3425	2.275	1	.131

Table 7.2: Generalised estimating equation results from study two.

When looking at whether the side the character was placed on effects whether participants choose to hit the person, there is a significance with participants more likely to hit characters that are on the right-hand side when under non-time pressure, compared to when they are on the left-hand side ($B = 0.573$, $p = 0.018$). This can be validated by looking at the cross-tab in figure 7.1.

It is clear from the results that the hypothesis is nullified as with the first study which also showed no significance between the independent and dependent variables. Whilst there is a significance between what side of the road a character is on and their likelihood of survival, this is not part of the research question.

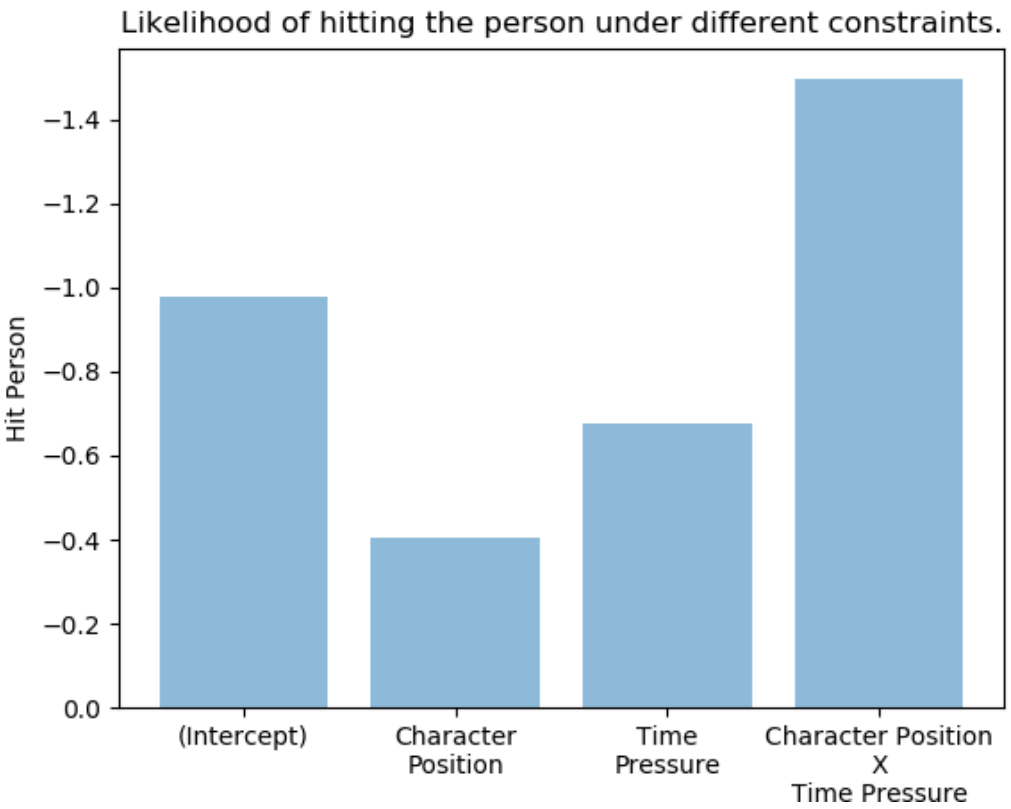

Figure 7.3: Parameters effects on participants likelihood to hit the person.

Participants First Attempts

During the study, it was observed that participants were likely to change their decision after the first attempt. To validate this, the same crosstab analysis was rerun on the first choice a participant made.

Was Time Pressure			Hit Person: Missed	Hit Person: Hit	Total
Non-Time	Character Position	Left	66.7%	33.3%	100.0%
		Right	30.0%	70.0%	100.0%
	Total		47.4%	52.6%	100.0%
Time	Character Position	Left	75.0%	25.0%	100.0%
		Right	33.3%	66.7%	100.0%
	Total		52.9%	47.1%	100.0%
Total	Character Position	Left	70.6%	29.4%	100.0%
		Right	31.6%	68.4%	100.0%
	Total		50.0%	50.0%	100.0%

Table 7.3: Crosstab results on first cases of participant decisions.

When looking at the crosstab in 7.3 compared to the previous one in 7.1, there is a change towards swerving from the left side, regardless of what object is in the way, whereas the prior analysis shows that participants are more likely to choose to save the person over the bollard. For this reason, an analysis was then carried out using GEE to identify any significance.

Parameter	B	Std. Error	Hypothesis Test: Wald Chi-Square	df	Sig.
(Intercept)	.770	.4940	2.431	1	.119
Was Time Pressure	.126	.7319	.029	1	.864
Character Position	.077	.4940	-.891	1	.876
Character Position X Was Time Pressure	-.280	.7319	.146	1	.702

Table 7.4: Generalised estimating equation results considering participants likelihood of going right.

As can be seen in 7.4 there is a preference for participants to swerve to the right-hand side on their first attempt, however the value derived is non-significant ($B = 0.770$, $p = 0.119$). It is difficult to conclude any form of significance, due to the sample size being reduced to forty choices instead of one-hundred and sixty. This data is

therefore being treated as only an interesting figure, and not one that should have conclusions drawn from it, unless further research was to be conducted.

Participants Last Three Attempts

Whilst the first attempt shows a preference to swerve to the right regardless of what is in the way, it is important to identify the effect the last three attempts had on the outcome. Should participants have been learning the study from the first attempt, the last three attempts may identify participants intended choices rather than sudden reactions.

Was Time Pressure			Hit Person: Missed	Hit	Total
Non-Time	Character Position	Left	87.1%	12.9%	100.0%
		Right	70.0%	30.0%	100.0%
	Total		78.7%	21.3%	100.0%
Time	Character Position	Left	67.7%	32.3%	100.0%
		Right	71.4%	28.6%	100.0%
	Total		69.5%	30.5%	100.0%
Total	Character Position	Left	77.4%	22.6%	100.0%
		Right	70.7%	29.3%	100.0%
	Total		74.2%	25.8%	100.0%

Table 7.5: Crosstab results on last three cases of participant decisions.

When looking at the crosstab in 7.5 compared to the previous one in 7.3 it is evident that there is a change in preference from swerving right, to avoiding the person in the majority of cases. It is clear that participants learned the study from their first attempt, and changed their response accordingly to what they deemed was correct. To identify if these values show a level of significance, a GEE analysis was conducted.

When looking at 7.6 it is possible to see there is a significant effect that participants in their last three attempts will try to avoid hitting the person ($B = -1.378$, $p = 0.000$). This, in contrast to the analysis of the first attempts, shows that participants change their decisions significantly once they understand the mechanics and outcomes of the environment. The effect of characters position causes a slight increase in the likelihood to hit the person, however the overall consensus is that participants will

Parameter	B	Std. Error	Hypothesis Test		
			Wald Chi-Square	df	Sig.
(Intercept)	-1.378	.3442	16.035	1	.000
Was Time Pressure	.549	.3705	2.199	1	.138
Character Position	.531	.3231	2.702	1	.100
Character Position X Was Time Pressure	-.618	.4092	2.283	1	.131

Table 7.6: Generalised estimating equation results considering participants likelihood of hitting the person.

still avoid hitting the person. Further to this, the value is non-significant so is only showing a random effect on participants decisions ($B = 0.531$, $p = 0.100$).

7.3.2 Qualitative

When considering the transcriptions that were produced from the study, three key themes were identified:

- Viewpoints about autonomous cars
- Moral decisions during the study
- The application of the decisions to real-life collisions

Within each of these, spawned multiple sub-themes which will be described below.

Viewpoints about Autonomous Cars

When viewing this key theme on a broad level, it is possible to see that positive and negative viewpoints are mixed, with no clear view being more present. To fully gain an understanding of what participants views were, it is best to delve into the key subthemes that were identified from the broad viewpoint theme.

Blaming human driving

The most common sub theme was participants blaming human driving for existing and future collisions, whether it be self-driving car related or not. Twelve participants

made reference to human driving being the cause of accidents, with one participant highlighting that:

> "the problem will still be people in charge of vehicles causing accidents that driverless cars struggle to avoid."

Other participants gave explanations for why this might be:

> "maybe humans don't cope very well with long periods of driving" and "you can't always trust a human to react as well as a machine in some instances".

With this identification of human error being the biggest factor for vehicle collisions, it does raise a joint outcome that participants are looking forward to autonomous cars to try to reduce these collisions with one participant stating that autonomous vehicles are

> "good things so long as everything is automated"

with another participant explaining the areas that they thought were of benefit:

> "forms of increased accuracy, better traffic management and just general replacing of humans in a specific role like taxi workers and transport"

This outcome is certainly the most interesting and reflects existing collisions that have occurred in the real-world, for example the woman in Arizona that was killed by an Uber which, according to BBC News, could have been avoided because a "police report suggests the car's driver was streaming an episode of talent show The Voice rather than monitoring the car's progress"(BBC News, 2018). This was emphasised by a participant who said:

> "I suppose that there is the danger that people will switch off in an automatic car isn't there?"

It is important to note though that due to the small sample of autonomous cars in the real-world, the chance of collisions between autonomous cars is far lower than with human life.

Diverting attention from driving

However, some people indicated that this was a feature they wanted from an autonomous car. A number of benefits were:

> "allowing people to sit there and do work whilst they're commuting"

which was also stated by another participant:

> "it might help me to do some work whilst being transported from one place to another" and attending to children meaning "you'd be able to pull your attention slightly and give it to them and just for those few seconds rather than spending two minutes pulling off and trying to deal with it".

One participant gave information that reflected another study where two out of five participants were willing to drive under the influence of alcohol in autonomous cars (Payre, Cestac and Delhomme, 2014, p. 255):

> "People that like to have a drink during the week, I think it will be beneficial for them".

However, only one participant mentioned this so doesn't match the two in five statistic of the aforementioned study.

Losing the ability to drive

Three participants expressed concern around autonomous cars potentially taking over the vehicle landscape due to the risk:

> "not from a safety perspective but from a driving enjoyment perspective."
> "I get enjoyment out of the motions of actually driving than just being automated so I'd lose that side of it."

With another participant further stating that:

> "little bit of me says it would be a very sad world if we didn't get to drive a car just cause I enjoy it".

For most of the participants, this was no generated as a concern for them, with some people saying that the loss of driving was welcome:

> "I would have no qualms about computers driving us instead. I think it's the future and it'll avoid a lot of accidents."

More Testing

Out of the forty participants, nine of them indicated concern around the need for more testing before they were willing to adopt such a vehicle:

> "They'd have to be used a lot before I chose to use one"

with another participant stating:

> "it would have to be proven to be absolutely safe before I was within the realm".

Not Being In Control of the Vehicle

The last sub-theme when considering participant viewpoints in relation to autonomous cars is participants being concerned with not being in control of the vehicle. There were eight participants that brought up issues around this, with some saying they would need to be fully in control:

> "I would have the fear of not being in control, even as a passenger, I'm not a good passenger I think I'd still need to be in control."

Whist others reported that they would require some ability to override the vehicle in an instance where they believed the vehicle wouldn't be able to handle a situation:

> "if you see danger and it you need to be able to have the ability to act before the car does and override what the cars doing. In case you see a danger and the car doesn't see a danger and you can override it."

Moral Decisions During the Study

This key theme is vital to the ability to answer the research questions at hand. Seven sub-themes were visible from the data collected.

Learning the Study

Due to the methodology utilising a within group design it was possible to see responses from participants where they were gradually changing their decisions over time. Ten out of the forty participants indicated levels of learning. The first example, one participant said:

> "I think I saw it better in time and after making my first two mistakes of hitting the person I just reacted a bit better cause I didn't want to hit him."

With another participant saying similar:

> "I'd learnt from my mistake, I'd learnt to hit the bollard and not the person."

Both of these examples present scenarios where participants felt they made mistakes in their decisions and so learnt to correct them after a couple of attempts.

Some participants actively decided, and then after a couple of attempts, chose to change their viewpoint:

> "I think I maybe just changed my opinion on it after running over the person twice, maybe I thought about it differently."

> "yeah you see this time I'm not sure, but my thoughts changed."

The two responses above were the occurrences of when someone actively changed their mind, but could not give an explanation as to why this was the case.

Looking for alternative options

A common sub-theme that emerged was participants looking for alternative options. This was visible in different skews, one being participants looking for more than two options:

> "I mean, I know it was set in there but of those choices on the second or third attempt I was looking, well can you get out, of the option AB is there an option C."

and those that would apply that logic to real life thinking:

> "in real life I probably would have gone up onto the pavement to avoid the barriers, I would have avoided all the obstacles if possible or basically brake, go slower."

The prior quote is one that is similar across a range of the quotes from this sub-theme. Many participants expected braking to be an option in the study, or felt that is what they would do in real-life compared to the Trolley Problem:

> "I would have turned my car side-ways to slow it down in time."

> "Obviously there would be a breaking situation."

> "I think it would rather stop than swerve."

> "Obviously if it was real life, I would have stopped but."

> "They'd be different because I'd put on the brakes."

This indicates that participants did not feel the Trolley Problem was how they envisaged an automated vehicle should behave.

Mistakes

One of the most common sub-themes, from twenty three participants, was stating that they had made a mistake when making their decisions.

The most common mistake was participants pressing the wrong button:

> "I was trying to avoid him but then I hit the wrong button. I think

because I saw him, and I saw him on that side my finger just kind of was like."

"Yep pressured the wrong bumper"

"Yeah, I think I might have clicked it a bit earlier or something to make that decision. I don't know really."

One of the other mistakes that arose was participants seeing the barrier before they saw the person meaning they reacted from this obstacle, putting them in the lane of the participant. Due to the way the study was designed, this meant they were unable to change the direction and found themselves hitting the person:

"At first I don't necessarily see the character saw a barrier, went and by the time I had seen the character it was kind of quite late and I'd already made those decisions and I don't know whether that was because couldn't see it or because of the speed came up quicker than I expected but that was, it was more of an obstacle in the road, and then seeing another obstacle means it was too late"

"Cause I just focused on the barrier."

Regret

Seventeen participants indicated feelings of regret from their decisions. All instances of regrets were in response to hitting a person:

"Regrets about the first two, I'd rather have gone into the bollard"

"The first one I was yeah cause as instant as I hit that person it just made me have a moment of feeling sick. I don't know why, it's just that I couldn't control it but then boom I've hit somebody."

"Maybe I should have missed the person in the first one, I was more focused on the bollard."

None of the participants indicated regrets about sacrificing themselves over hitting the character in the road. This can be validated by one of the participants:

"Yeah, I don't regret smashing a car or injuring myself so that's my choice as for the other person it's not their choice to be there."

Trusting Safety Features

When participants chose to sacrifice themselves by hitting the bollard, there were twelve participants that trusted the safety features of the vehicle to protect them in comparison to hitting the person where they felt the safety features of a car hitting a pedestrian were lower.

"I would take that the car would have air bags and would save me and the risk of hurting me is far less than hitting a person with a car"

"I think in the split second I would steer myself into an obstacle and I'd trust the safety systems in my vehicle to protect me, rather than deliberately driving into somebody"

"I would hopefully in a real-life situation would aim to do. I've got more chance of avoiding injury with airbags and crumple zones and bumpers and bla-de-bla-de-bla"

This demonstrates that the participants assumptions of vehicles are more reliable than that of a traditional trolley.

Unwillingness to Hurt Someone

The following two sub-themes are the opposite of one another; the benefit of this is being able to break the understanding of why someone would or wouldn't wish to hurt another person.

Predominantly participants chose to sacrifice themselves over another person. Thirty-five participants chose to sacrifice themselves, either all of the time or occasionally.

There were similar reasons as to why this was the case, with some participants stating it was the morally correct thing to do:

"Yeah, my thinking there was, well I'm in a car and he's not and so I'll

> probably do less affluent things, I hope. It would do less damage to me than him and that was the moral choice"

Some participants also valued life over "inanimate objects" and didn't wish to hurt them:

> "I would always choose the inanimate object."

> "Because I would never want to hit somebody with my car that would kill them than when I hit a bollard and it would be pretty safe."

This outcome does raise questions. Although participants were informed by the researcher that in the scenario, a collision with the bollard would result in the participants death, the quotes above do indicate that this may have not become apparent to a large number of participants, potentially explaining the heavy weighting towards self-sacrifice.

Willingness to Hurt Someone

In the opposite situation where participants chose to hit the character, ten participants actively hit the character. In these instances, participants were explaining their reasoning for killing the person with the arguments around comparison of the quantity saved and quantity killed:

> "It's a one to one so one person survives one person doesn't survive. Yeah so, it's just one to one so it doesn't really, maybe there's no, there's no like disadvantage, its either you hit a thing and you die that person survives or you hit a guy, he dies, and you survive so yeah."

A common occurrence was that of participants accusing the character of being in the road, thus justifying their decision:

> "I'm not going to crash into a barrier and kill myself if some bloke is stood in the middle of the road."

> "I saw a barrier and thought, well, he's going to be stood there so I thought I'd aim for the person."

Application of Gameplay to Real-Life Moral Dilemmas

The final key theme is participants applying the virtual reality environment to real-life moral dilemmas. For this key theme, there were three sub-themes identified, which were categories of what participants could say.

Changing Decision Compared to Gameplay

Twelve participants stated that they would be likely to change their decisions compared to the gameplay. It is important to note that this does not mean those participants did not state the opposite, only indicating that there were instances that they felt they would.

> "In real life I probably would have gone up onto the pavement to avoid the barriers, I would have avoided all the obstacles if possible or basically brake, go slower. Don't put yourself in that situation."

> "I would drive much more slowly, and I would feel that I would have the opportunity to stop or to see what was coming and take some evasive action."

> "So, I saw that pedestrian a lot longer before I actually hit them, so I would have put on the brakes and been able to slow down."

Once again, braking is a clear option that participants felt was required in order to make decisions that they felt were relevant to them.

There was also a participant that wanted to change their decision due to making a mistake:

> "I would hope they'd differ, so I wouldn't hit the person, but I mean if you're under a pressure I don't suppose it's a split-second decision. But yeah, I'd hope they'd differ."

Keeping Decision Similar to Gameplay

Alternatively, there were twenty-eight participants that stated they would keep some or all of their decisions the same.

> "I'd like to think not, I'd like to think that if that was the choice I'd still make the choice that I'm in a car and I'd probably come off and it's, I'd be responsible for that vehicle in a sense"
>
> "They'd be similar, yeah. As I said, yeah human lives aren't replaceable."
>
> "I think they would be pretty much the same cause as I said, because I would always try and avoid hitting anyone with my car, because I'm the one driving it so, so I should be paying attention to the things around me rather than them having to pay attention to what I'm doing so I would feel like it's my responsibility rather than theirs."

Unsure on the Decision

The final sub-theme relates to where participants were unsure what their decision would be in real-life. There were nine participants that stated they were unsure.

> "I wish I could answer that question I would like to think that yes, I would make the decision but in reality, would I, I don't know. I'm being honest, I knew that wasn't real."
>
> "I don't really have a way of telling what I would decide in a split-second cause it could depend on as many factors as it's possible to have."

7.4 Conclusion

This chapter has indicated that participants, when in a within-group design, are likely to respond in a self-sacrificial way, to ensure those outside the vehicle are unharmed. This was presented by both qualitative and quantitative results. The quantitative result showed a significant effect that participants are likely to swerve towards the bollard, regardless of time-pressure. The qualitative provided a detailed

explanation for why this might be, showing that participants would prefer to "choose the inanimate object."

Alongside the additional detail of why participants would sacrifice themselves, the qualitative data presented other areas of interest. One, most notable, being the participants decisions to try and find alternative options than the Trolley Problem's binary choice format. This is elaborated on further within the discussion section.

Results also indicated an unexpected outcome. Participants in their first simulation compared to future attempts are more likely to swerve to the right hand side. This is regardless of what is in front of them. Whilst non-significant, the results are still interesting and are further explained in the discussion section.

However, as stated in the qualitative results, there is question as to whether participants understood the outcome that they would die should they hit the bollard. This is evident from quotes such as

> "I would take that the car would have air bags and would save me and the risk of hurting me is far less than hitting a person with a car."

Therefore, it could be viewed that the assumption the vehicle would not hurt them could be present. This would not provide a one-to-one comparison of life or death. This instead would show a life to slightly hurt comparison, which does not equate. Thus, results should be treated with caution with the knowledge this was present.

8. Discussion

Both studies indicate no correlation between time and non-time pressure. The first study also indicated that there was no correlation between placing participants in different areas of the environment and having an effect on the decisions made. It is therefore clear that the hypothesis for both studies are nullified.

8.1 Comparing the Moral Machine to this Study

Study one, whilst having the hypothesis nullified, does provide interesting insight into the decisions made. Firstly, participants were more likely to save human characters over animals. This was a significant difference from one another, with cats being the least likely to cause participants to swerve away from. This is compared to young characters who cause the highest weighting. These results are comparable to the Moral Machine's which showed that participants are more likely to spare young children, with animals being the most likely to be sacrificed in a collision scenario. See figure 5.1 about the Moral Machine's findings.

Due to the actor independent variable being found to be non-significant, participants did not change their decision according to their position in the environment. Further information of why this could be is explained in the limitations section.

The implication of the above arguments is that the results gathered from the Moral Machine are more likely to be valid when implementing ethical algorithms centred around the Trolley Problem. This is because the time pressured results show no significant difference compared to non-time pressured. This indicates that the approach the Moral Machine took was valid in gaining participants decisions. However, this can be argued against due to the sample sizes between the Moral Machine and these studies being different. However, the study was conducted in isolation of the Moral

Machine evaluating both independent variables, thus checking for any significance within its own primary data. It is therefore possible to imply, having found similar viewpoints, that the Moral Machine's results are valid.

However, where the validity of both this study and the Moral Machine's remains, is that pedestrians would be around vehicles for some considerable time. Removing humans from the driving equation does not guarantee an autonomous vehicle freedom from human intervention. For an autonomous vehicle to not need to have any form of ethical or collision algorithms installed, would be when humans are removed from the vehicle's environment entirely for example through bridges or subways. This would allow pedestrians to travel without interfering on autonomous vehicles. Achieving this would be when this research is less likely to be of concern to the general public.

An argument derived from the literature review, is that there is a lack of anti-robot functionality in the Moral Machine. Whilst this is still the case, and there is no evidence to neither prove or disprove the interference of robots on the results, there is an argument that despite this, the results of the web survey line up with the Moral Machine. Due to the web survey being protected by anti-robot functionality, it can be safely argued that whatever effect robots could have had on the Moral Machine can be seen as less of an issue. This is due to validation of the web survey on the social decisions being made.

8.2 Self-Preservation vs Self-Sacrifice

When looking at the results of the second study via quantitative analysis, participants are more likely to sacrifice themselves over saving themselves. This is also justified during the audio interviews where many participants stated they would rather sacrifice themselves than knowingly hurt someone else. Further evidence of this is present in the first study where, regardless of independent variable, participants were likely to select the bollard. This outcome is in keeping with another study where participants were presented with the option of killing themselves or two others. 52% of the time they chose themselves. Interestingly, in the aforementioned

study, the more characters that were added into the road, the more self-sacrificial a participant became, with the result of seven people in the road achieving 70% self-sacrifice rate (Bergmann et al., 2018, p. 6). In comparison, this study showed participants were more likely to avoid the person when it was a one-to-one comparison. What the study does not show, is the effect when an increase of character count occurred. Had more characters been added, a similar increase may have been observed. This effect was shown in the first study, indicating that standard females, and young and old males, caused a significant effect on causing participants to be more likely to select the bollard.

The unexpected result from study two is that participants were more likely to hit the person when they are on the right-hand side with no time pressure. This is unexpected because the participants are more likely to sacrifice themselves over others; why non-time pressure then causes participants to slightly favour avoiding the bollard on the right hand-side, is somewhat complex to answer. There was no indication from the qualitative data that indicated any change in thought process. There is also the argument that it could be random chance due to everything else being visible in the GEE as insignificant.

8.3 Instinctual vs. Moral

Study two also shows two interesting perspectives of how participants react to a collision scenario. Firstly, when looking at the initial decision that participants made, crosstabulation indicated that participants, regardless of what is on the road, are more likely to swerve to avoid what is in front of them. Compare this to future attempts, where participants are more likely to avoid hitting the person, regardless of what side they are on. The reason why participants are more likely to avoid what's in front of them in the first scenario, appears to be classed as a mistake from participants, mainly when they hit the person. This is evident by the qualitative data gathered where participants stated the following:

> "I think I saw it better in time and after making my first two mistakes of

hitting the person I just reacted a bit better cause I didn't want to hit him."

Further to this in the quantitative section, an analysis was carried out on the last three choices, which identified that participants, regardless of character location, would be significantly inclined to avoid hitting the person and sacrificing themselves.

Further evidence of this can be seen in the results section.

When most participants chose to sacrifice themselves over hitting the person, they felt more comfortable with the decision and stated they would not change their opinion should they do it again.

This raises an interesting implication for the development of ethical collision systems in autonomous vehicles. From the results, there could be two implementations, the instinctual and the moral. Developing an instinctual ethical system, based on the results above, would indicate that the vehicle should be fifty percent likely to swerve from whatever is in front of it. In contrast, the moral implementation would be more willing to choose the route where external human damage is at a minimum, compared to that of inside the vehicle. Arguably, the implementation of an instinctual method would need more research and revision in order to be deemed as a viable option due to the small sample size available.

8.4 Humans are the Issue, Not the Machine

When looking at the qualitative data of study two, it is evident that from the forty participants who took part in the study, twelve indicated feelings around humans being the main cause for vehicle collisions. This is in keeping with some autonomous vehicle collisions that have occurred (for example the woman in Arizona that was killed by an uber which, could have been avoided because a "police report suggests the car's driver was streaming an episode of talent show The Voice rather than monitoring the car's progress" (BBC News, 2018)). Whilst it is arguable that this is in fact the vehicle's fault for not identifying the person, it is important to note

the vehicle is not running under level five autonomy so was relying on the driver's alertness to take over when the vehicle wasn't able to react accordingly.

If these were to be factored into the argument of developing ethical collision systems, it would be possible to argue that the benefit of developing them would be an interim one. Whilst human drivers are present on the road, the ability for an autonomous vehicle to handle unexpected situations and collisions are required. In contrast, if all vehicles were to become autonomous, the implementation of ethical algorithms would be less of a concern due to there being lower occurrences of situations where a human could cause disruption to the vehicle's driving procedure.

8.5 Ineffectiveness of the Trolley Problem in Real-Life Dilemmas

Another area highlighted by study two, is the inability for the Trolley Problem to account for true real-life dilemmas in vehicles. Many participants stated feelings around wanting more options to select from than just choosing a binary option. Several of them stated that braking would have been one of their main choices rather than swerving.

> "they'd be different because I'd put on the brakes."

> "obviously if it was real life, I would have stopped but."

This joins up with the arguments provided by Goodall who states that "an automated vehicle needs a way to determine if the benefits of moving into the left lane outweigh the costs" (Goodall, 2016, p. 815). The argument posed in the literature review against this is the need to understand how people feel when a collision was to occur. The results gathered indicate that participants do not see the Trolley Problem as a viable solution for autonomous vehicle collision systems and need to be more dynamic, accounting for wider factors than a binary decision.

The implications of this align with Reese's argument that it is not his idea to "give engineers basic material for programming a self-driving car's moral decisions" (Reese,

2016). Should the Trolley Problem be implemented as an ethical algorithm, it would limit the scope to what is an already dynamic environment.

One alternative to the Trolley Problem that may alleviate the issues, is the Tunnel Problem. This would encase the vehicle in a tunnel, meaning participants options became limited. However, this would only remove some of the variables. In reality, vehicles are not at a constant speed or going in a straight line. Vehicles can slow down, deviate from their course and this is just a small example. One of the main thoughts about the tunnel problem is that the participants may believe they can use the side of the tunnel as a method to slow down in time, or believe they can use the brakes. The main way to rectify this is by providing strict rules about what the vehicle can do, but this sacrifices the ability to compare to real life circumstances.

The Trolley Problem does have its strengths. It allows decisions to be narrowed down to comparable options, making conclusions far easier to identify as well as "understanding real-world reactions" (Collins, 2018). What this means for autonomous vehicles is that decisions within the Trolley Problem are a finite glance into what people believe is ethically correct. Understanding this, means autonomous cars can react correctly in Trolley Problem dilemmas and provide future vehicle owners with more trust that their vehicle will react in a manor they believe is correct.

Overall, the Trolley Problem is just a snippet into the understanding of ethical dilemmas in vehicles. Alternative methods should be used in the future to identify any differences in decisions from participants and gain a broader understanding of the dynamic environment that is the driving task.

8.6 The Observance of Time-Constraint Affecting Bollard Decisions

In the web study, there was a significant difference between participants likelihood to hit the bollard when they were under the effect of time pressure, with almost double the effect of hitting the bollard compared to non-time pressure. It is important to

mention that this observation is unlikely to be true due to significance having a 5% chance of giving false positives (Colquhoun, 2017, p. 1). Whilst it may appear as significant, the majority of results were non-significant and having only a few significant values does raise concerns that these false positives are being displayed.

8.7 Limitations

This course of research had limitations, largely meaning there are more areas to explore in the future. Firstly, the web survey, although it had a relatively large sample size, was not comparative to that of the MIT Moral Machine's. Had it been of that scale, it could have been possible to compare the data between the two. This would allow some interesting information to be gathered. However, by gaining results for time and non-time pressure, data could be compared in isolation to this study alone, so the lack of comparison to the Moral Machine is only a minor inconvenience.

A similar argument can be made for the second study. Quantitatively the sample size was smaller than desired. It is not possible to infer whether the viewpoints gathered reflect the Lincolnshire populace as well as populace on an international level which would be argued as impossible to interpret with the current sample. However, qualitatively, the number of participants gathered is enough to gain thematic evidence which provides additional insight into the quantitative information.

Another area of the first study that had limitations, was its isomorphic design. Having followed a similar design choice to the MIT's Moral Machine, the actor independent variable was put into the description above the images rather than visually within the image itself. This could be a reason for why the results are derived as non-significant for the actor independent variable. It is therefore important to conclude that the use of the actor independent variable was only tested using textual format and so the results could vary if someone was to research the effect when the person's location was visually shown in the environment.

Another weakness of the visual design was that of the characters in the vehicle during the bollard images. Because only the heads were shown and were not as visually

prominent as those on the road, there could be a bias that causes participants to focus more on who is in front of the vehicle, compared to who is in the vehicle. This would contradict the assumption that life both in the vehicle and on the road is of an equal weighting.

The first study has another limitation around the length of time participants had to prepare themselves and then the double time-to-collision that has been previously reported, as the average does not account for different values of time-to-collision and instead focuses on a liberal value. The results cannot therefore be used to assess different levels of time-pressure.

For study one, when considering the bollard images, due to the already complex dataset, it was not possible to analyse the data, taking into account the position of the characters on the road. The reason for this was due to the comparison factor being a bollard, thus it was not possible to create a difference value between the character types, unlike that of the standard image scenarios.

Whilst the second study used a randomised option set to assign to the different collision sets, there is a known issue that due to using a random number generator from .Net, the decision sets generated are not truly random. This caused the first and last decision sets to be used more in the last collision scenario that a participant would undertake. Due to using within-group design this is not an issue because participants still completed all possible collision sets in a non-human controlled assignment. This only becomes an issue when looking across the decisions made rather than as a collective. For example when looking at the Instinctual vs. Moral argument presented earlier. This is further elaborated within the future studies section, which provides advice for how to prove or disprove the findings of the VR study.

The web survey was unable to stop participants from completing the study multiple times; this was a trade-off to ensuring the study was completely anonymous. Any attempt to trace who had completed the study would have provided some form of personal identifier which was not requested as part of the ethical application. This means the survey could have been influenced by several participants that do not reflect that of the results. This however is less of a concern because the MIT Moral

Machine followed the same approach. Further to this argument, the randomisation of scenarios and independent variables provided an incredibly small chance that participants would see the same scenario, and a one in six chance of being in the same sample group.

8.8 Recommendations

8.8.1 Future Studies

An area for further research, would be if time and non-time pressure situations have any difference in decision times. It could be possible that participants make decisions in similar time frames regardless of circumstance, which could negate any argument for time-pressure causing a difference to participants choices. This data gathering was not conducted as part of this research due to the richness of data being collected. For a single researcher the quantity of data gathered was difficult to analyse, providing more would have made data analysis even more difficult within the time constraints of the research project. The argument against this form of research is asking what it would contribute to. The knowledge from the research is that participants are more likely to make the same decision. What would the knowledge that participants choosing a decision in a different time window contribute? If a significant difference was to be identified, it would be ideal to combine this with gaining an insight into participants level of comfort and stress levels in each independent variable. This would identify if the independent variables do have a positive or negative impact on participants decisions, regardless of the fact they are likely to make the same decision. Answering the question of what it would contribute can now be answered; the knowledge of if time and non-time pressure has an impact on decision time and stress levels, would indicate how the ethical and automotive communities should target their future research and developments.

Further to the prior recommendation, research around different levels of time pressure should be considered. This is due to the limitation of the double time-to-collision

value which could have provided a more liberal decision time to that of real-life collision scenarios. Should the results show that this continues to make little difference on participants decisions, it can be concluded that time-pressure has no effect on the outcomes chosen. On the contrary to this, should the results show a significance the more time pressure is applied, then further questions should be asked into what this could mean for the understanding of the results found in both this study and prior studies, such as the Moral Machine.

Another area that requires more investigation, is identifying whether the sample size was the cause of the first scenarios of the second study emerging as non-significant where participants were showing signs that they were more likely to swerve from whatever was in front of them, regardless of obstacle. Should this be discovered to be the case, the questions around participants learning the second study would be further validated and instead could raise questions around whether what the Moral Machine and the web survey of this thesis, provides a voting-based decision where participants choose the outcome they prefer, compared to an instinctual decision which the first decision of the second study may hint towards. Once again, this study does not determine that this is true based on it being insignificant, however the sample size raises doubts that what's being observed could be proven by a more direct study.

8.8.2 Practical Actions

It is important to argue that creating an implementable ethical algorithm is not ideal when using both the Moral Machine and the results of this thesis. The main reason is discrimination, as outlined by the Federal Ministry of Transportation and Digital Infrastructure that "In the event of unavoidable accident situations, any distinction based on personal features (age, gender, physical or mental constitution) is strictly prohibited. It is also prohibited to offset victims against one another." (Federal Ministry of Transport and Digital Infrastructure, 2017, p. 11). Creating a system that can distinguish between race leads to a dangerous area that can be abused and targeted. Therefore, collision algorithms should follow that of the Federal Ministry:

"the protection of human life enjoys top priority in a balancing of legally protected interests. Thus, within the constraints of what is technologically feasible, the systems must be programmed to accept damage to animals or property in a conflict if this means that personal injury can be prevented"(Federal Ministry of Transport and Digital Infrastructure, 2017, p. 11).

From these points and previous expression in this document, it is not appropriate to say the Trolley Problem is a suitable model for collision decisions in automated vehicles. The possibilities of collision scenarios are far broader than the Trolley Problem can achieve. However, the Trolley Problem is a useful method of gaining focused, human responses to ethical decisions which would otherwise be difficult to collect when evaluating real-life collision possibilities. The results from this thesis should therefore be viewed as Human viewpoints rather than technical guidelines or advice to the implementation of collision logic.

8.9 Conclusion

This research aimed to identify if there was a significant difference between time and non-time pressure, as well as significant differences on participant's decisions based on where they were in the environment during a Trolley Problem modelled collision scenario. Based on the quantitative results from the web survey and the quantitative and qualitative results of the VR study, it can be concluded that there is no significance difference in either of these hypothesis. The results indicate that participants are likely to respond in a utilitarian manor, regardless of the two independent variables used.

Due to the research attempting to identify the effect time-pressure had on participant's decisions, there was an expectation that the effect time-pressure would have would become apparent from the results, with the implication being that research would be required into understanding the moral implications between time and non-time pressure, thus applying that knowledge to autonomous vehicles. However, the results showed very little effect on participants outcomes, further validating prior

results of research such as that of the Moral Machine's. The expectation of this was because a similar design pattern and analysis method was used to easily line up the two outcomes side-by-side as well as prior research by Sutfeld who found that the lowering of time pressure, caused "tendency toward social desirability" and "would likely rely on slower cognitive processes, and thus not come into effect in fast-paced intuitive decisions" (Sütfeld et al., 2017, p. 10). This effect was not observed in this research. This could have been caused by one of the prior mentioned limitations; due to the time-to-collision being double that of the average values found, potentially causing less time-pressure being induced than expected. Further to this, there was no recorded value of decision times across both time and non-time pressure. However, this research does show that during these time-to-collision scenarios, participants are likely to make similar decisions to that of when no time pressure is involved. Further research should therefore be considered by testing the methodology used across a different range of time pressures, ignoring the actor independent variable.

Whilst the hypothesis' were nullified, the methodologies used were valid. The randomisation of the scenarios in the web survey and the iterative assignment of participant groups did mean that decisions were made on random datasets, avoiding any bias. Further to this, the use of five seconds of count down and then three seconds of decision time, ensured participants had time to prepare for the option, before seeing it, as well as the web browser having time to render all of the elements. There is the argument that this time to prepare and the double average time-to-collision value used could have caused less time-pressure than what would be desirable, however ensuring that participants do get a chance to make a choice rather than their browser be the cause for not making a decision, was deemed as more important. As recommended in the discussion section, research could be carried out into the effect different time pressure values have on participant's decisions, which would either validate, or disprove the findings found in this research.

The use of virtual reality was beneficial in removing the effect browser render time may have on participant's decisions, as well as attempting to induce a more immersive experience to that of the web survey. It also provided new insights into why participants made their decisions which other studies had not fully gathered; provid-

ing understanding that in a life or death situation, participants are more likely to sacrifice themselves to make a morally acceptable choice.

This research has provided further validation and understanding into existing Trolley Problem dilemmas, most notably the Moral Machine, due to the results of both the web survey and the VR study indicating similar character preference profiles, the contributions also indicate that certain time-pressures and the location of a character in an environment do not have an effect on the choices made by participants.

However, the research has also argued and shown the limitations of the Trolley Problem when utilised in vehicle collisions, evidently shown through the VR study, highlighting the weakness of the binary choice model. This was later recommended that vehicle manufacturers should avoid using the Trolley Problem data as suitable for training the autonomous vehicles which require further development into social understanding during driving tasks, as well as collision tasks which the results in this and prior studies of this nature would not be able to provide.

References

Aebersold, K. (2019). *Software Testing Methodologies.* URL: https://smartbear.com/learn/automated-testing/software-testing-methodologies/ (visited on 8th Feb. 2018) (cit. on pp. 25, 48).

Awad, Edmond et al. (2018). 'The moral machine experiment'. In: *Nature* 563.7729, p. 59 (cit. on pp. 2, 11, 41).

Barribal, L. K. and A. While (1994). 'Collection data using semi-structured interview: a discussion paper.' In: *Journal of Advanced Nursing* 19, pp. 328–335 (cit. on p. 57).

Basso, A. and M. Miraglia (2008). 'Avoiding Massive Automated Voting in Internet Polls.' In: *Electronic Notes in Theoretical Computer Science* 197.2, pp. 149–157. URL: https://www.sciencedirect.com/science/article/pii/S1571066108000637 (visited on 10th Dec. 1018) (cit. on p. 12).

BBC News (2018). 'Arizona Uber crash driver was 'watching TV'.' In: URL: https://www.bbc.co.uk/news/technology-44574290 (visited on 28th June 2019) (cit. on pp. 66, 80).

Bergmann, Lasse T et al. (2018). 'Autonomous Vehicles Require Socio-Political Acceptance—An Empirical and Philosophical Perspective on the Problem of Moral Decision Making'. In: *Frontiers in behavioral neuroscience* 12, p. 31 (cit. on pp. 15, 16, 79).

Brey, Philip (1999). 'The ethics of representation and action in virtual reality'. In: *Ethics and Information technology* 1.1, pp. 5–14 (cit. on p. 16).

Brown, Barry (2017). 'The social life of autonomous cars'. In: *Computer* 2, pp. 92–96 (cit. on pp. 8, 9).

Collins, J. (2018). *Hypotheticals versus the real world: The trolley problem [blog].* URL: https://jasoncollins.blog/2018/07/04/hypotheticals-versus-the-real-world-the-trolley-problem/ (visited on 12th May 2020) (cit. on p. 82).

Colquhoun, D. (2017). 'The reproducibility of research and the misinterpretation of p-values'. In: *R Soc Open Sci* 4.12 (cit. on p. 83).

Contissa, Giuseppe, Francesca Lagioia and Giovanni Sartor (2017). 'The Ethical Knob: ethically-customisable automated vehicles and the law'. In: *Artificial Intelligence and Law* 25.3, pp. 365–378 (cit. on p. 20).

Crawford, Kate and Ryan Calo (2016). 'There is a blind spot in AI research'. In: *Nature News* 538.7625, p. 311 (cit. on pp. 18, 19).

Cristofari, Cécile and Matthieu J Guitton (2014). 'Surviving at any cost: guilt expression following extreme ethical conflicts in a virtual setting'. In: *PloS one* 9.7, e101711 (cit. on pp. 7, 8).

Dixit, Avinash and Barry Nalebuff (2019). *Prisoners' Dilemma*. The Library of Economics and Liberty. URL: https://www.econlib.org/library/Enc/PrisonersDilemma.html (visited on 6th July 2019) (cit. on p. 20).

D'Olimpio, Laura (2016). 'The trolley dilemma: would you kill one person to save five?' In: URL: http://theconversation.com/the-trolley-dilemma-would-you-kill-one-person-to-save-five-57111 (visited on 5th July 2019) (cit. on p. 3).

Easterbrook, S. (2001). *Lecture 12: Software Design Quality [lecture]*. URL: http://www.cs.toronto.edu/~sme/CSC444F/slides/L12-SoftwareQuality.pdf (visited on 8th Feb. 2019) (cit. on pp. 27, 49).

Federal Ministry of Transport and Digital Infrastructure (2017). *Ethics Commission Automated and Connected Driving*. Berlin. URL: https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile (visited on 18th May 2018) (cit. on pp. 19, 86, 87).

Foot, P. (1967). 'The Problem of Abortion and the Doctrine of the Double Effect'. In: 5, pp. 1–6. URL: Available%20from%20https://philpapers.org/archive/FOOTPO-2.pdf (visited on 19th Jan. 2019) (cit. on p. 9).

Fournier, Tom (2016). 'Will my next car be a libertarian or a utilitarian?: Who will decide?' In: *IEEE Technology and Society Magazine* 35.2, pp. 40–45 (cit. on p. 19).

Friedman, Doron et al. (2014). 'A method for generating an illusion of backwards time travel using immersive virtual reality—an exploratory study'. In: *Frontiers in psychology* 5, p. 943 (cit. on p. 14).

Gibbs, S. (2018). 'Uber's self-driving car saw the pedestrian but didn't swerve – report'. In: URL: https://www.theguardian.com/technology/2018/may/08/ubers-self-driving-car-saw-the-pedestrian-but-didnt-swerve-report (visited on 17th June 2018) (cit. on p. 17).

Gogoll, Jan and Julian F Müller (2017). 'Autonomous cars: in favor of a mandatory ethics setting'. In: *Science and engineering ethics* 23.3, pp. 681–700 (cit. on pp. 20, 21).

Goodall, Noah J (2016). 'Away from trolley problems and toward risk management'. In: *Applied Artificial Intelligence* 30.8, pp. 810–821 (cit. on pp. 22, 81).

Graham, K. (2017). *Vehicle-to-vehicle communication tech going nowhere [blog]*. URL: http://www.digitaljournal.com/tech-and-science/technology/vehicle-to-vehicle-communication-tech-going-nowhere/article/503010 (visited on 3rd May 2019) (cit. on p. 23).

Hao, K. (2018). 'Should a self-driving car kill the baby or the grandma? Depends on where you're from'. In: URL: https://www.technologyreview.com/s/612341/a-global-ethics-study-aims-to-help-ai-solve-the-self-driving-trolley-problem/ (visited on 22nd Feb. 2019) (cit. on p. 2).

Hevelke, Alexander and Julian Nida-Rümelin (2015). 'Responsibility for crashes of autonomous vehicles: an ethical analysis'. In: *Science and engineering ethics* 21.3, pp. 619–630 (cit. on p. 21).

Honda Government Relations (2014). *Honda Demonstrates Advanced V2P and V2M Safety Technologies [video]*. URL: https://vimeo.com/73407713 (visited on 3rd May 2019) (cit. on p. 23).

Hong, J. and K. Ottoboni (2017a). *Generalized Estimating Equations (GEE) [blog]*. URL: https://rlbarter.github.io/Practical-Statistics/2017/05/10/generalized-estimating-equations-gee/ (visited on 3rd May 2020) (cit. on p. 40).

– (2017b). *Generalized Estimating Equations (GEE) [blog]*. URL: https://rlbarter.github.io/Practical-Statistics/2017/05/10/generalized-estimating-equations-gee/ (visited on 15th June 2019) (cit. on p. 40).

House of Commons Library (2017). *Connected and autonomous road vehicles*. London. URL: http://researchbriefings.files.parliament.uk/documents/CBP-7965/CBP-7965.pdf (visited on 19th May 2018) (cit. on pp. 6, 20).

Hox, Joop J and Hennie R Boeije (2005). 'Data collection, primary versus secondary'. In: (cit. on pp. 33, 54).

Jardine Motors Group (2019). *The History of Car Technology*. Jardine Motors Group. URL: https://news.jardinemotors.co.uk/lifestyle/the-history-of-car-technology (visited on 22nd Feb. 2019) (cit. on p. 2).

Jurdak, R. and S. S. Kanhere (2018). 'Who's to blame when driverless cars have an accident?' In: URL: http://theconversation.com/whos-to-blame-when-driverless-cars-have-an-accident-93132 (visited on 22nd Feb. 2019) (cit. on p. 2).

Karpov, Alexander (2017). 'Preference diversity orderings'. In: *Group Decision and Negotiation* 26.4, pp. 753–774 (cit. on p. 38).

Knapman, C. (2016). 'How long until we have fully driverless cars?' In: URL: https://www.telegraph.co.uk/cars/features/how-long-until-we-have-fully-driverless-cars/ (visited on 22nd Feb. 2019) (cit. on p. 2).

Knight, W. (2015). *Car-to-Car Communication [blog]*. URL: https://www.technologyreview.com/s/534981/car-to-car-communication/ (visited on 3rd May 2019) (cit. on p. 22).

Körber, Moritz, Eva Baseler and Klaus Bengler (2018). 'Introduction matters: Manipulating trust in automation and reliance in automated driving'. In: *Applied ergonomics* 66, pp. 18–31 (cit. on p. 7).

Kusano, D. K. and H. Gabler (2011). 'Method for Estimating Time to Collision at Braking in Real-World, Lead Vehicle Stopped Rear-End Crashes for Use in Pre-Crash System Design'. In: *Passenger Card - Mechanical Systems* 4.1, pp. 435–443. DOI: https://doi.org/10.4271/2011-01-0576. URL: https://doi.org/10.4271/2011-01-0576 (cit. on p. 34).

Lanteri, Alessandro, Chiara Chelini and Salvatore Rizzello (2008). 'An experimental investigation of emotions and reasoning in the trolley problem'. In: *Journal of Business Ethics* 83.4, pp. 789–804 (cit. on pp. 12, 13).

Levin, S. and J. C. Wong (2018). 'Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian'. In: URL: https://www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe (visited on 20th May 2018) (cit. on p. 17).

Lin, Patrick (2016). 'Why ethics matters for autonomous cars'. In: *Autonomous driving*. Ed. by B. Lenz M. Maurer J. C. Gerdes and H. Winner. Berlin, Germany: Springer, Berlin, Heidelberg, pp. 69–85 (cit. on p. 21).

Microsoft (2015). *C# Coding Conventions (C# Programming Guide)*. Microsoft. URL: https://docs.microsoft.com/en-us/dotnet/csharp/programming-guide/inside-a-program/coding-conventions (visited on 1st Mar. 2019) (cit. on p. 27).

Millar, J. (2014). *An ethical dilemma: When robot cars must kill, who should pick the victim? [blog]*. URL: https://robohub.org/an-ethical-dilemma-when-robot-cars-must-kill-who-should-pick-the-victim/ (visited on 3rd May 2019) (cit. on p. 16).

Nyholm, Sven (2018). 'The ethics of crashes with self-driving cars: A roadmap, I'. In: *Philosophy Compass* 13.7, pp. 1–10 (cit. on p. 12).

Ohn-Bar, Eshed and Mohan Manubhai Trivedi (2016). 'Looking at humans in the age of self-driving and highly automated vehicles'. In: *IEEE Transactions on Intelligent Vehicles* 1.1, pp. 90–104 (cit. on pp. 6, 18).

Parsons, Thomas D (2015). 'Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences'. In: *Frontiers in human neuroscience* 9, p. 660. URL: https://www.frontiersin.org/article/10.3389/fnhum.2015.00660 (visited on 8th Dec. 2018) (cit. on p. 16).

Patil, I. et al. (2014). 'Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas'. In: *Soc Neurosci* 9.1, pp. 94–107 (cit. on p. 14).

Payre, William, Julien Cestac and Patricia Delhomme (2014). 'Intention to use a fully automated car: Attitudes and a priori acceptability'. In: *Transportation research part F: traffic psychology and behaviour* 27, pp. 252–263 (cit. on pp. 6, 7, 67).

Philips, Brian and Tom Morton (2015). *Making Driving Simulators More Useful for Behavioral Research—Simulator Characteristics Comparison and Model-Based Transformation*. Version FHWA-HRT-15-016. US. Department of Transportation, pp. 1–26. URL: https://www.fhwa.dot.gov/publications/research/ear/15016/15016.pdf (visited on 6th July 2019) (cit. on p. 7).

Poushter, J. (2016). *Internet access growing worldwide but remains higher in advanced economies*. URL: http://www.pewglobal.org/2016/02/22/internet-access-growing-worldwide-but-remains-higher-in-advanced-economies/ (visited on 21st July 2018) (cit. on p. 36).

Reese, J. (2016). 'MIT's 'Moral Machine' crowdsources decisions about autonomous driving, but experts call it misguided'. In: URL: https://www.techrepublic.com/article/mits-moral-machine-crowdsources-decisions-about-autonomous-driving-but-experts-call-it-misguided/ (visited on 10th Dec. 2018) (cit. on pp. 11, 81).

Renda, Andrea (2018). 'Ethics, algorithms and self-driving cars–a CSI of the 'trolley problem''. In: *CEPS Policy Insight* 2018/02. URL: https://www.ceps.eu/system/files/PI%202018-02_Renda_TrolleyProblem.pdf (visited on 20th May 2018) (cit. on p. 17).

Schreurs, Miranda A and Sibyl D Steuwer (2015). 'Autonomous driving-political, legal, social, and sustainability dimensions'. In: *Autonomous Driving*. Ed. by B. Lenz Maurer J. C. Gerdes and H. Winner. Berlin, Germany: Springer, pp. 149–171 (cit. on p. 21).

Skulmowski, Alexander et al. (2014). 'Forced-choice decision-making in modified trolley dilemma situations: a virtual reality and eye tracking study'. In: *Frontiers in behavioral neuroscience* 8, pp. 1–16 (cit. on pp. 14, 15).

Suarez, A. (2018). *How and why our experiments with virtual reality motion made us ill [blog]*. URL: https://venturebeat.com/2018/02/27/how-and-why-our-experiments-with-virtual-reality-motion-made-us-ill/ (visited on 22nd Feb. 2019) (cit. on p. 49).

Sütfeld, Leon R et al. (2017). 'Using virtual reality to assess ethical decisions in road traffic scenarios: applicability of value-of-life-based models and influences of time pressure'. In: *Frontiers in behavioral neuroscience* 11, pp. 1–13 (cit. on pp. 15, 88).

Technative (2018). *AI Ethics and The Tunnel Problem [blog]*. URL: https://www.technative.io/ai-ethics-and-the-tunnel-problem/ (visited on 3rd May 2019) (cit. on p. 16).

The Tesla Team (2015). *Your Autopilot has arrived [blog]*. URL: https://www.tesla.com/en_AU/blog/your-autopilot-has-arrived (visited on 22nd Feb. 2019) (cit. on p. 2).

Thomson, J. J. (1985). 'The Trolley Problem'. In: *The Yale Law Journal* 94.6, pp. 1395–1415. URL: http://www.oswego.edu/~delancey/trolley.pdf (visited on 8th June 2018) (cit. on pp. 3, 9, 10).

Vaughn, G. (2017). *Should Your Next Survey be Anonymous or Not? [blog]*. URL: https://blog.zef.fi/en/should-your-next-survey-be-anonymous-or-not (visited on 15th Dec. 2018) (cit. on p. 37).

Wiegmann, Alex, Yasmina Okan and Jonas Nagel (2012). 'Order effects in moral judgment'. In: *Philosophical Psychology* 25.6, pp. 813–836. DOI: 10.1080/09515089.2011.631995. eprint: https://doi.org/10.1080/09515089.2011.631995. URL: https://doi.org/10.1080/09515089.2011.631995 (cit. on p. 13).

Yadron, D. and D. Tynan (2016). 'Tesla driver dies in first fatal crash while using autopilot mode'. In: URL: https://www.theguardian.com/technology/2016/jun/30/tesla-autopilot-death-self-driving-car-elon-musk (visited on 17th June 2018) (cit. on p. 17).

Appendix

Parameter		B	Std. Error	Hypothesis Test Wald Chi-Square	df	Sig.
	(Intercept)	-.484	.1836	6.947	1	.008
	Green Light	.019	.2055	.008	1	.927
	Driver	-.071	.2275	.098	1	.754
	Bystander	-.038	.2391	.025	1	.873
Time	Green Light	-.449	.2315	3.754	1	.053
	Cat	.224	.2057	1.185	1	.276
	Dog	.099	.2405	.169	1	.681
	Young Female	-.090	.3027	.088	1	.767
	Old Female	.236	.2794	.711	1	.399
Driver	Standard Female	.406	.3368	1.452	1	.228
	Young Male	.407	.3302	1.516	1	.218
	Old Male	.186	.2662	.487	1	.485
	Standard Male	-.059	.3624	.026	1	.871
	Green Light	.182	.2709	.451	1	.502
	Cat	.477	.1905	6.262	1	.012
	Dog	.437	.2068	4.470	1	.034
	Young Female	.065	.3023	-.046	1	.829
	Old Female	-.191	.2793	.468	1	.494
Passenger	Standard Female	.225	.3570	.396	1	.529
	Young Male	.102	.3741	.074	1	.786
	Old Male	.137	.3214	.181	1	.671
	Standard Male	.188	.4096	.212	1	.645
	Green Light	.145	.2862	.258	1	.611

Table 8.1: Table showing all non-bollard Generalised Estimation Equation results from the web survey that are not included in results section. The autonomous actor type is the base Intercept so has not been included.

Parameter		B	Std. Error	Hypothesis Test Wald Chi-Square	df	Sig.
	(Intercept)	.526	.2131	6.092	1	.014
	Time	.512	.2048	6.240	1	.012
Time	Bollard	-.694	.1590	19.036	1	.000
	Cat	-.122	.1491	.668	1	.414
	Dog	.111	.1428	.600	1	.439
	Young Female	.049	.1542	.102	1	.749
	Old Female	-.120	.1524	.615	1	.433
	Standard Female	-.364	.1722	4.464	1	.035
	Young Male	-.238	.1581	2.259	1	.133
	Old Male	-.047	.1614	.084	1	.772
	Standard Male	-.380	.2513	2.284	1	.131
	Autonomous	.040	.2440	.027	1	.869
	Passenger	.057	.2567	.049	1	.825
Autonomous	Bollard	-.059	.1855	.102	1	.750
	Cat	.108	.1736	.388	1	.534
	Dog	.044	.1751	.062	1	.804
	Young Female	.274	.1895	2.084	1	.149
	Old Female	.196	.1949	1.009	1	.315
	Standard Female	-.015	.2039	.006	1	.940
	Young Male	.119	.1892	.397	1	.529
	Old Male	.096	.1899	-.276	1	.612
	Standard Male	.119	.3481	.117	1	.732
Passenger	Bollard	.069	.1789	.148	1	.700
	Cat	.008	.2070	.001	1	.970
	Dog	.323	.1789	3.257	1	.071
	Young Female	.254	.1806	1.977	1	.160
	Old Female	.181	.1783	1.035	1	.309
	Standard Female	.147	.2121	.483	1	.487
	Young Male	.393	.1874	4.396	1	.036
	Old Male	-.302	.2031	-.700	1	.137
	Standard Male	.509	.3409	2.226	1	.136

Table 8.2: Table showing all bollard Generalised Estimation Equation results from the web survey that are not included in results section. The driver actor type is the base Intercept so has not been included.